

## From source code to software as a service, how to make software more accessible?

deRSE 23, February 2023

Uwe Konrad, Tobias Huste (HZDR)

Uwe Jandt (DESY)

David Schäfer, Thomas Schnicke (UFZ)

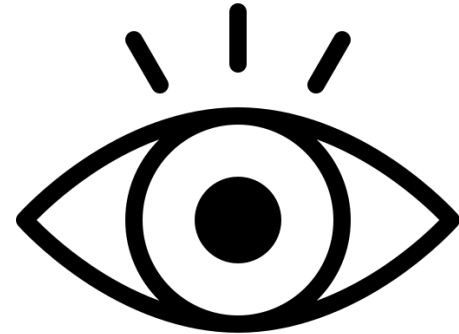
DOI: [10.5281/zenodo.7653477](https://doi.org/10.5281/zenodo.7653477)

# Motivation

---

Modern Software is used as a Service (SaaS) in the cloud, also in science.  
What are the **advantages**?

- The software is **visible and accessible** and supports the work in large teams.
- It **is scalable** and enables the use of large storage and computing resources.
- It offers **high availability, trustworthiness** and latest releases.
- SaaS platforms **support workflows** connecting different applications.
- **Usually help** and support is provided



But what are **implications** for the software development and engineering?

- The talk addresses these questions.
- The illustrative examples relate to the [helmholtz.cloud](https://helmholtz.cloud) platform.

# Approaches of a software development

There are **different approaches** to software development

- A SW solution for a specific **project** or a research task
- A SW **product** is to be developed for a market.
- A SW **service** is to be created for a cloud **platform**.

The approaches define certain **conditions** for a project, illustrated by a maturity diagram.

A **SW service**:

- can achieve **increased visibility and recognition**
- requires **continuous funding** and a larger team
- high quality: **mature**, reproducible development and engineering **processes**
- Active user community, open source: **incremental, distributed development**

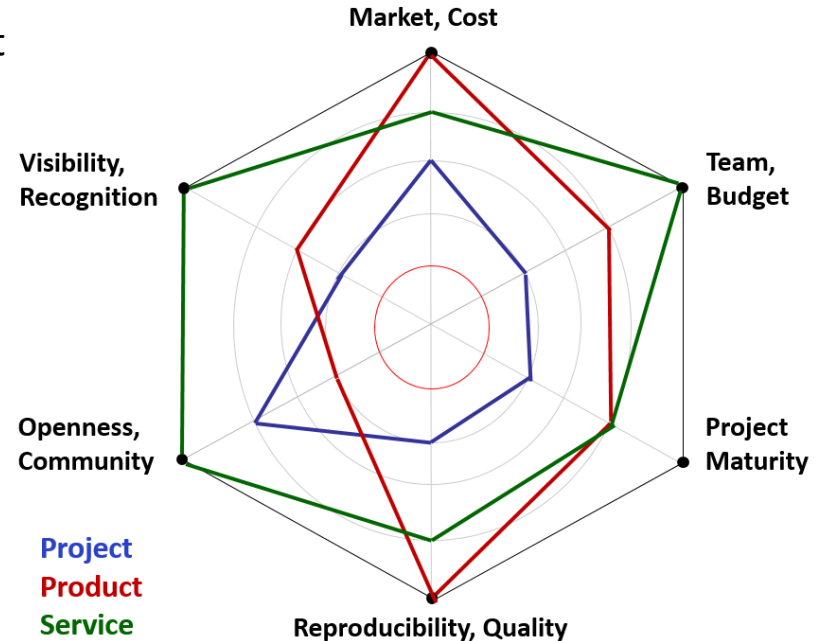


Fig. 1: Exemplary maturity diagram

# The Software Development Lifecycle

Specific features of SaaS in the software development lifecycle:

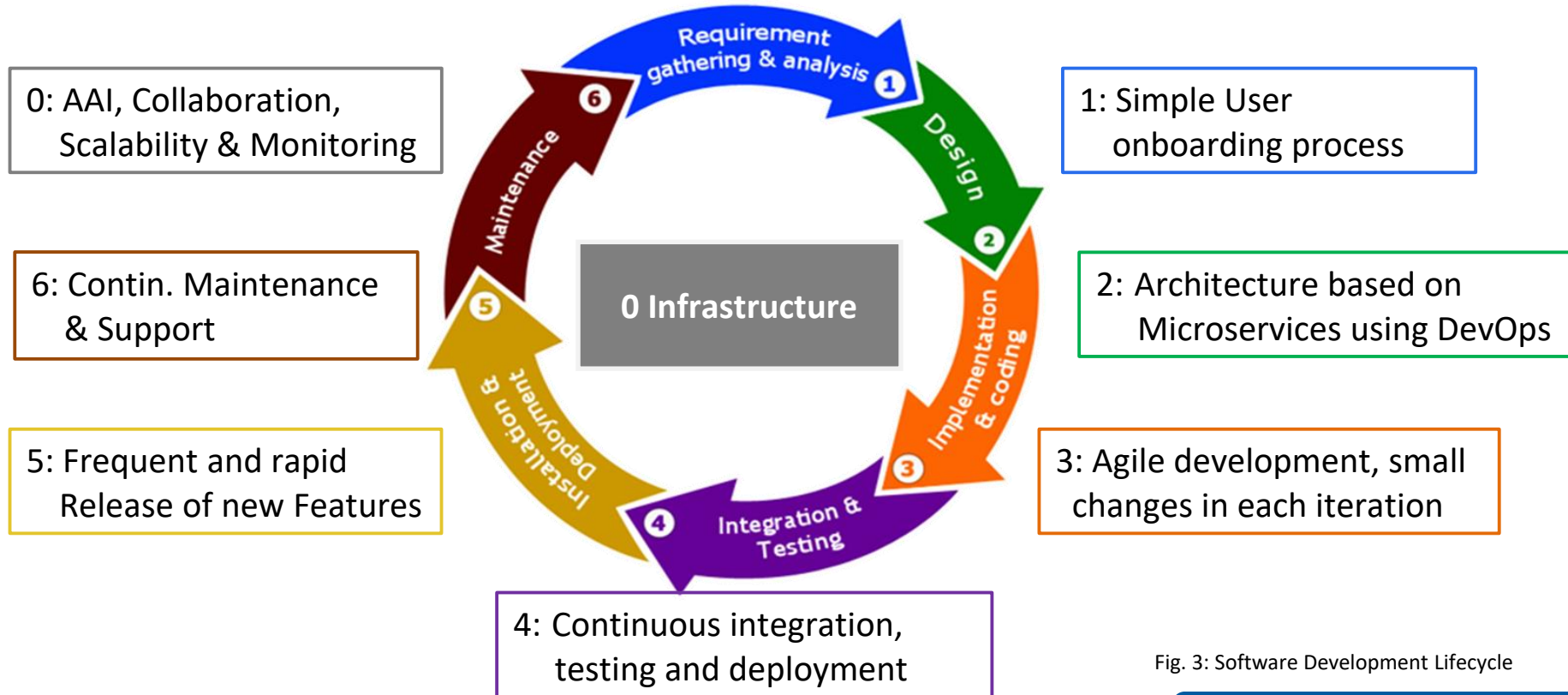
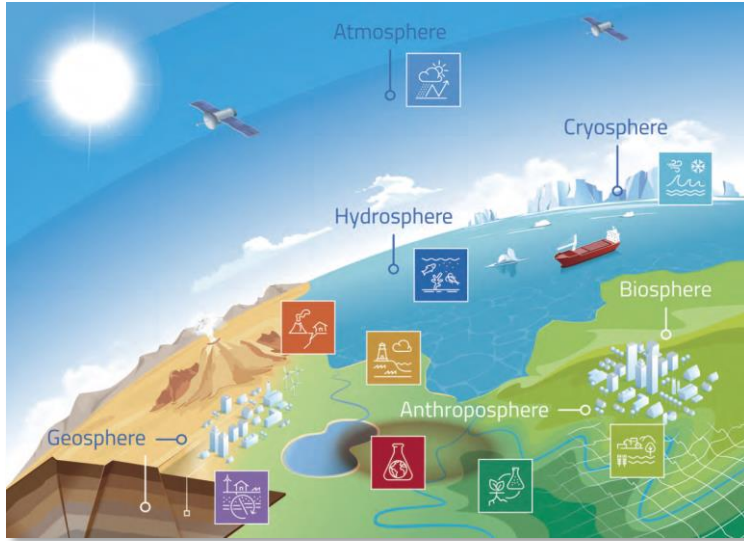


Fig. 3: Software Development Lifecycle



-> Establish an interoperable digital ecosystem for Earth System Science supporting the holistic research perspective

- Digitalization of data acquisition, workflows and pipelines address digital needs in science
- Interaction with incubator activities and use of services
- Guiding principles are Open Science and FAIR Data
- Use cases defined within the program research
- Implementation provides added value for research and the public
- Connection to national and international data infrastructure, e.g. NFDI and EOSC



Fig. 4: Helmholtz Datahub Earth-and-Environment  
see [DataHub](#) and [Helmholtz RF Earth and Environment](#) websites

## What is SaQC?

- A Python module, a command line application and, increasingly, a Software Service
- An framework to quality control of timeseries data
- A collection of **methods and algorithms** to:
  - process (resampling, interpolation, smoothing)
  - quality control (breaks-, constants-, outlier-, noise-, drift-detection) and
  - annotate **timeseries data**

## SaQC as Software Development Project:

- Small team of 3 to 5 part time SW developers and data scientists from the  **UFZ** HELMHOLTZ Centre for Environmental Research
- Initial commit 03.2019, first 'stable' release in 11.2021
- Open Source, License GPL 3

<https://git.ufz.de/rdm-software/saqc>

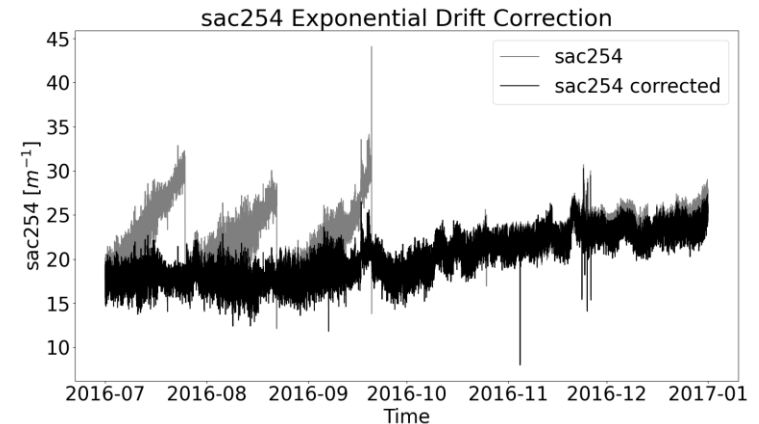
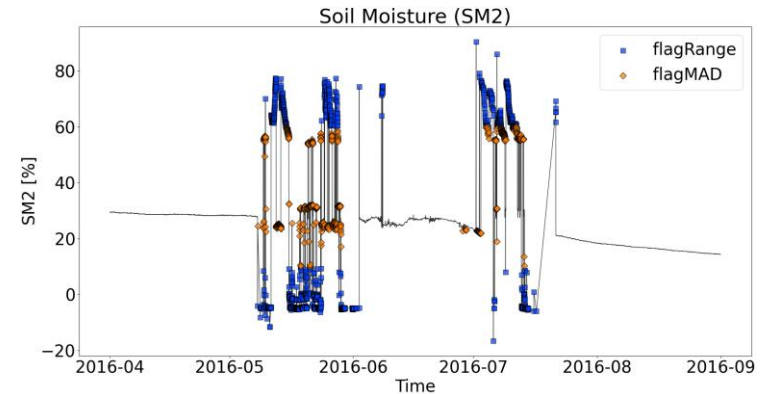


Fig. 5: SaQC on-line documentation

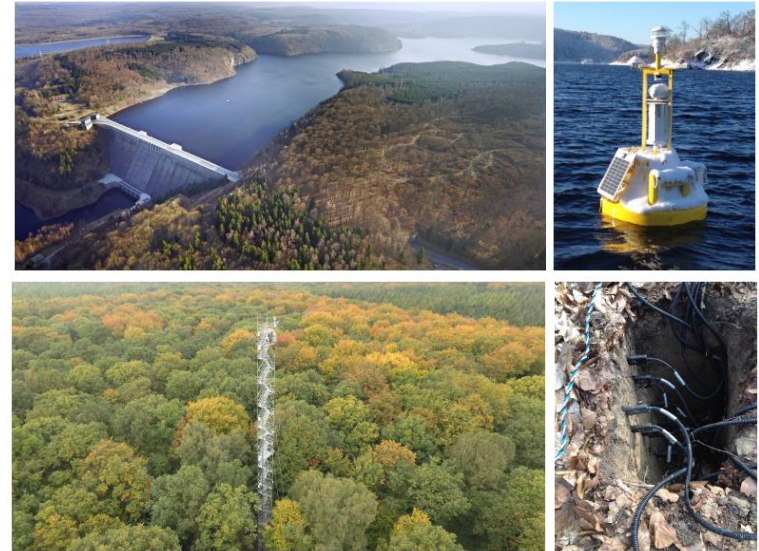
## What do we use it for?

- Quality control of timeseries data from terrestrial observatories
- As part of fully automatized data workflows from sensors into databases

## Use cases of TERENO

- [Bode Hydrological Observatory](#)
  - 912 Sensors, 42 Dataloggers
  - > 185 mio. observations and growing
- [Intensive Test Site Hohes Holz](#)
  - 2114 sensors, 87 loggers
  - > 775 mio. data values and growing

**TERENO**  
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES



[Fig. 6: tereno.net](#)

## The long way from product to SaaS

- 2019: Source distributed, locally deployed CLI Application
- 2020: Package distributed Python API
- 2022: [Configuration App](#) as an centrally hosted, graphical user interface
- 2022: SaQC is [Helmholtz Software Spotlight](#)
- 2023: GitLab CI/CD deployed quality control pipelines:
  - Reduce the entry barrier for non/less technical staff
  - Build pipelines to run project specific SaQC setups on cloud infrastructure
- 2024: Planned integration into a fully-fledged Timeseries Management System:
  - Deploy SaQC as a service within a larger scale system to **manage data flows** from source to sink
  - Integrate SaQC into an ecosystem of community research data management tools



## How does a Cloud Platform (HIFIS) help with that?

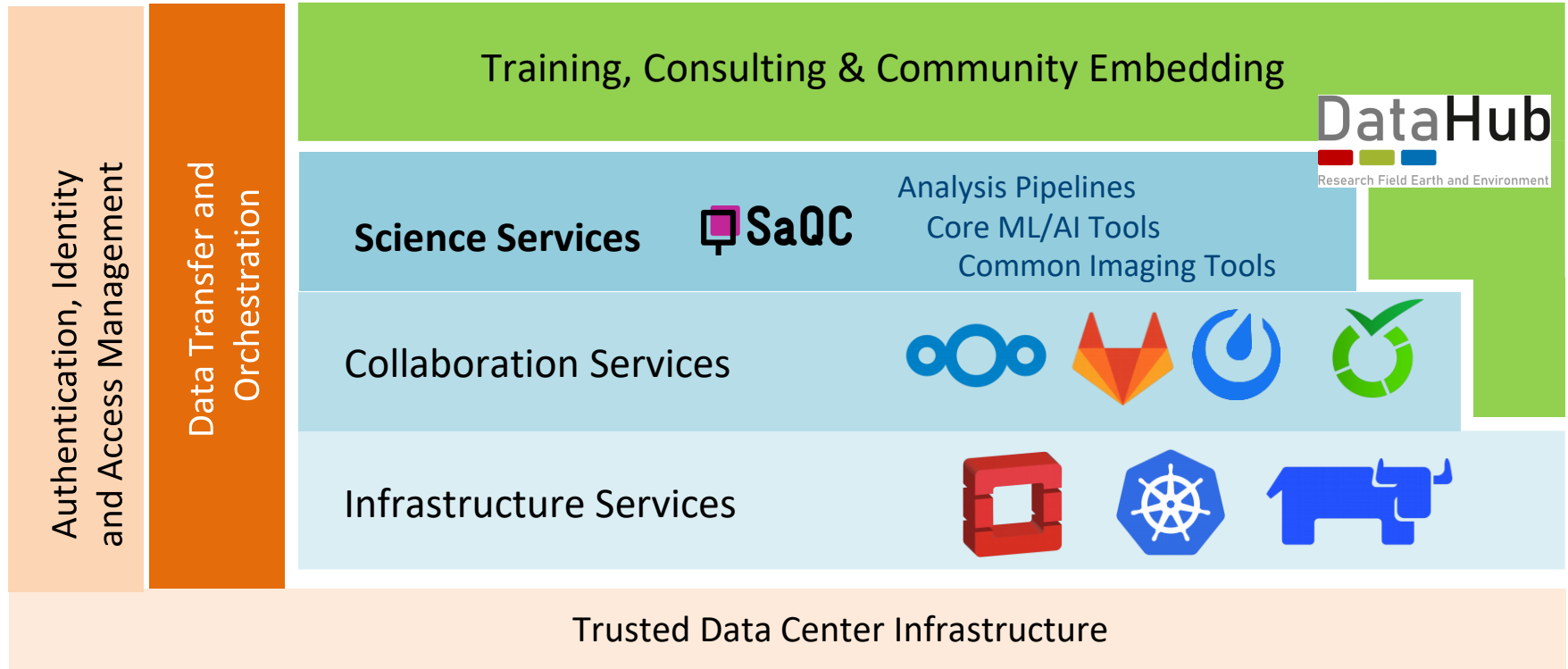


Fig. 7: [hifis.net](https://hifis.net)

## Helmholtz Codebase Architecture

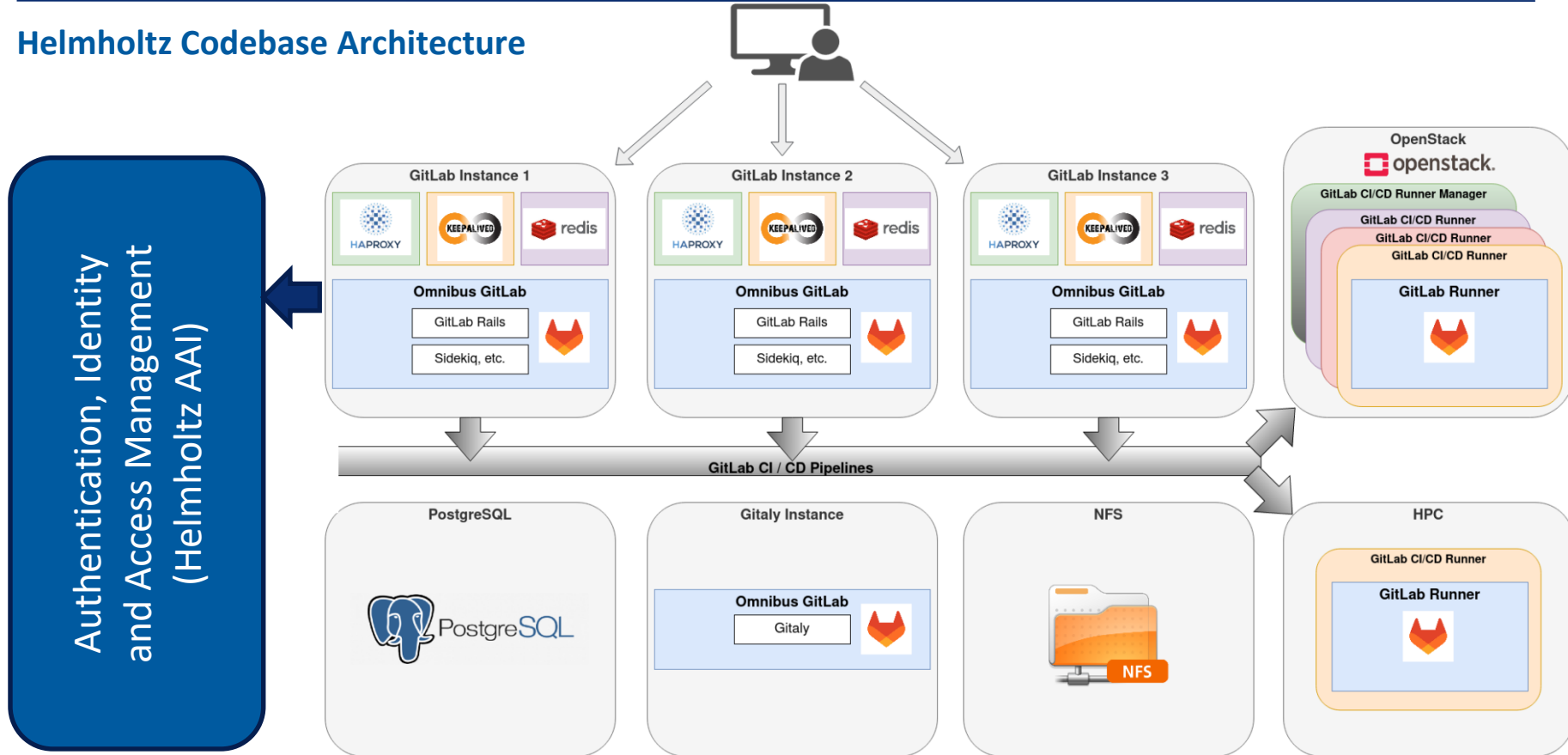
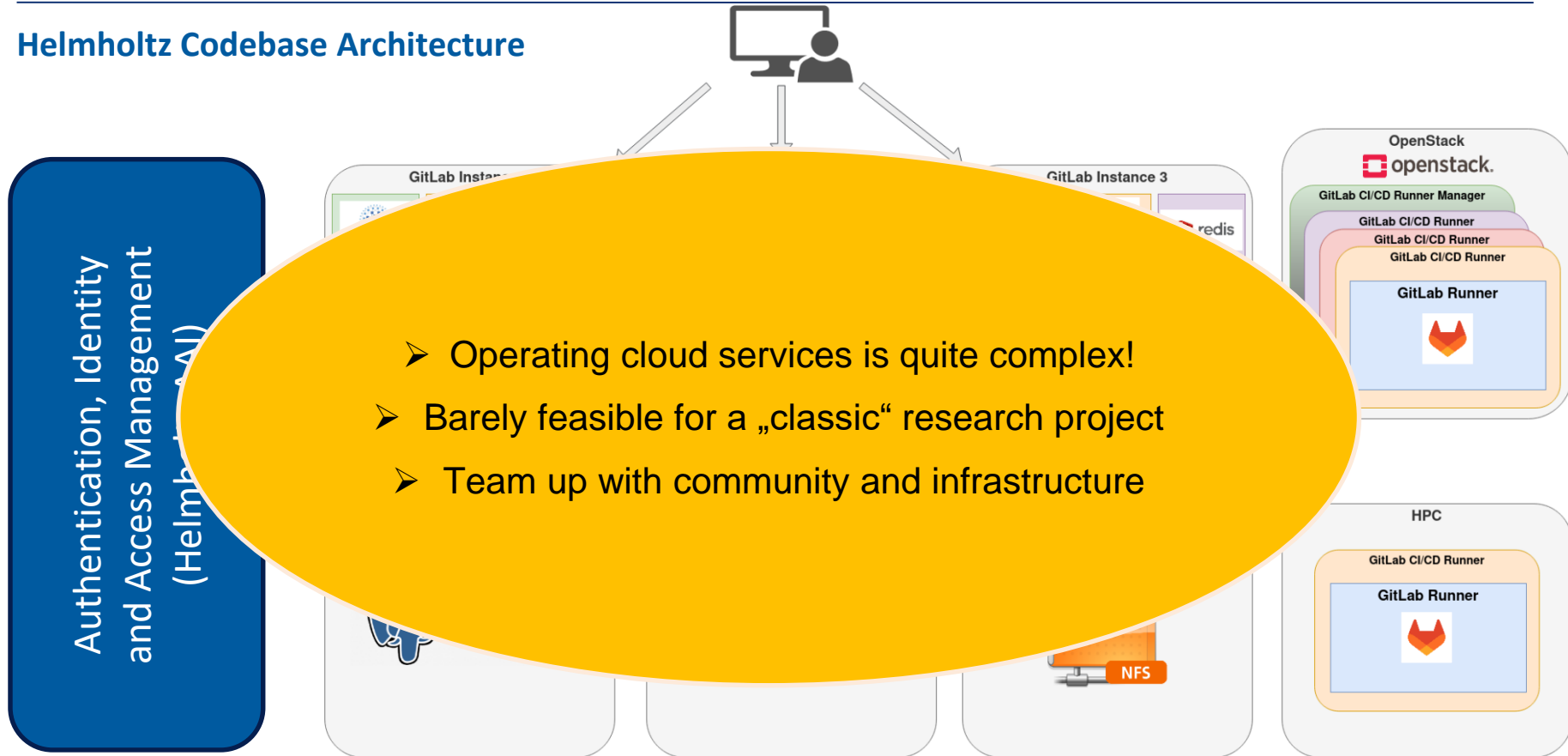


Fig. 8: [hifis.net](https://hifis.net)

## Helmholtz Codebase Architecture



Authentication, Identity  
and Access Management  
(Helmholtz IAM)

Fig. 8: [hifis.net](https://hifis.net)

## Mastering complexity through using re-usable and generic solutions

- Building services on top of community standards, e.g. Ansible, HELM
- Provide reusable service provisioning recipes:  
<https://galaxy.ansible.com/hifis>, <https://github.com/hifis-net>



HashiCorp  
**Terraform**

## Allow scientists to build their own portable cloud-ready applications which can be embedded in research workflows!

- Testdrive community driven applications on federal cloud platforms
- Scientific communities provide test-instances and usable applications
- a SW platform can be a **market place of ideas** and solutions



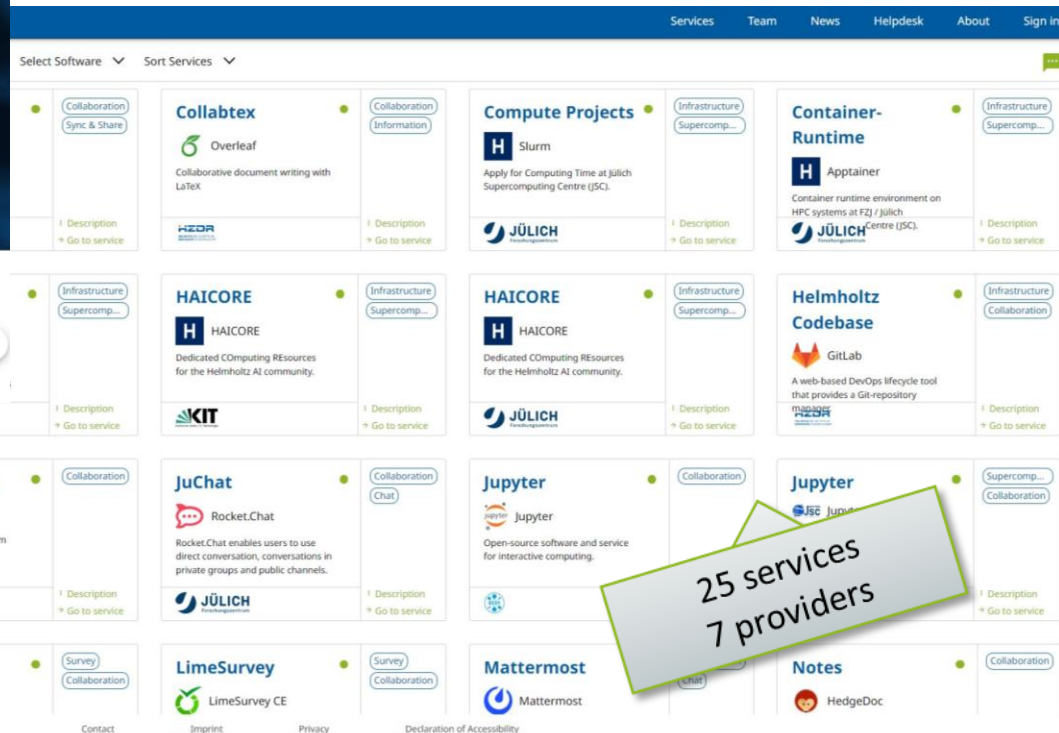
## Helmholtz Cloud

In the Helmholtz Cloud, members of the Helmholtz Association of German research centers provide selected IT-Services for joint use.

Note: Helmholtz Cloud is currently in pilot phase and will go in production during 2022.

[BROWSE ALL SERVICES](#)

Find more information at [hifis.net](https://hifis.net)



The screenshot shows the HIFIS service catalog interface. At the top, there is a navigation bar with links for Services, Team, News, Helpdesk, About, and Sign in. Below the navigation bar, there are filters for 'Select Software' and 'Sort Services'. The main content area displays a grid of service cards, each representing a different service. Each card includes the service name, a brief description, and a 'Go to service' link. The services shown include:

- Collabtex**: Collaborative document writing with LaTeX. Provider: HEZDR.
- Compute Projects**: Apply for Computing Time at Jülich Supercomputing Centre (JSC). Provider: JÜLICH.
- Container-Runtime**: Container runtime environment on HPC systems at FZJ / Jülich. Provider: JÜLICH.
- HAICORE**: Dedicated Computing Resources for the Helmholtz AI community. Provider: KIT.
- Helmholtz Codebase**: A web-based DevOps lifecycle tool that provides a Git-repository. Provider: JÜLICH.
- HIFIS Events**: An Events Management service for use within Helmholtz and their partners, based on Indico. Provider: JÜLICH.
- HIFIS Helpdesk**: HIFIS Helpdesk Ticketing System based on Zammad. Provider: HEZDR.
- JuChat**: Rocket.Chat enables users to use direct conversation, conversations in private groups and public channels. Provider: JÜLICH.
- Jupyter**: Open-source software and service for interactive computing. Provider: JÜLICH.
- Jupyter on HAICORE**: Jupyter on HAICORE. Provider: HAICORE.
- LimeSurvey**: LimeSurvey CE. Provider: HEZDR.
- Mattermost**: Mattermost. Provider: JÜLICH.
- Notes**: HedgeDoc. Provider: HEZDR.

At the bottom of the page, there are links for Contact, Imprint, Privacy, and Declaration of Accessibility.

5 services categories  
strong team  
powerful infrastructure

25 services  
7 providers

Fig. 9: helmholz.cloud

## Facts & Figures

>13.000

individual users in AAI

>2.000

non-Helmholtz users

All Centres

One Login using Helmholtz AAI!



Helmholtz Digital  
Services for  
Science  
—  
Collaboration  
made easy.

>65

collaborating groups (VO)

27

active cloud services  
of 8 providers

5

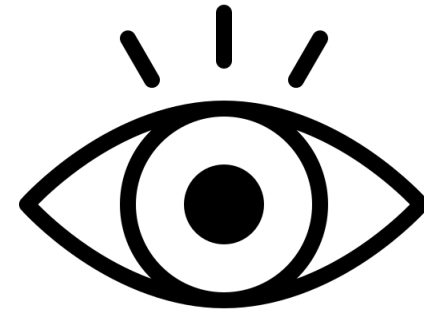
service categories:  
Infrastructure, Storage,  
HPC, Collaboration and  
Science

>900

processed support  
tickets in 2022

## Providing Software as a Service on a Research Cloud Platforms helps to

- enable a new quality of collaborative, data-centred research
- achieve greater impact and sustainability
- deploy cutting-edge technology based on a new technological paradigm



### Visit us @:

- <https://hifis.net>
- <https://helmholtz.cloud>
- <https://datahub.erde-und-umwelt.de>
- <https://helmholtz.software/software/sagc>



[Eye icons created by Freepik - Flaticon](#)