



**HIFIS**

HELMHOLTZ  
FEDERATED  
IT SERVICES

**HELMHOLTZAI** | ARTIFICIAL INTELLIGENCE  
COOPERATION UNIT

## Helmholtz Federated IT & Accessible Compute Ressources for Applied AI Research

Uwe Jandt (DESY) on behalf of  
HIFIS, HAICORE, and Helmholtz AI initiatives

Nov 02, 2020

# The Six Research Fields of Helmholtz

# HELMHOLTZ

RESEARCH FOR  
GRAND CHALLENGES

ENERGY



EARTH AND  
ENVIRONMENT



HEALTH



AERONAUTICS,  
SPACE AND  
TRANSPORT



MATTER

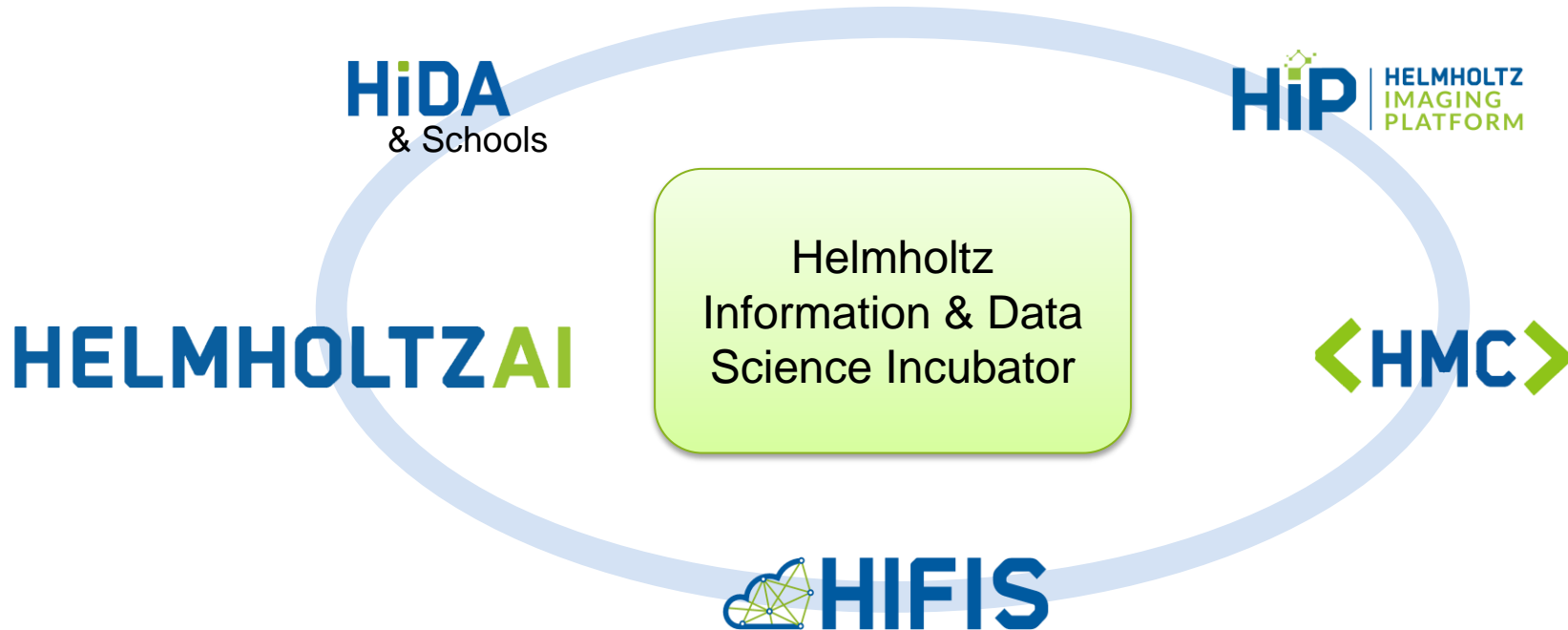


KEY  
TECHNOLOGIES  
(IN THE FUTURE:  
INFORMATION)



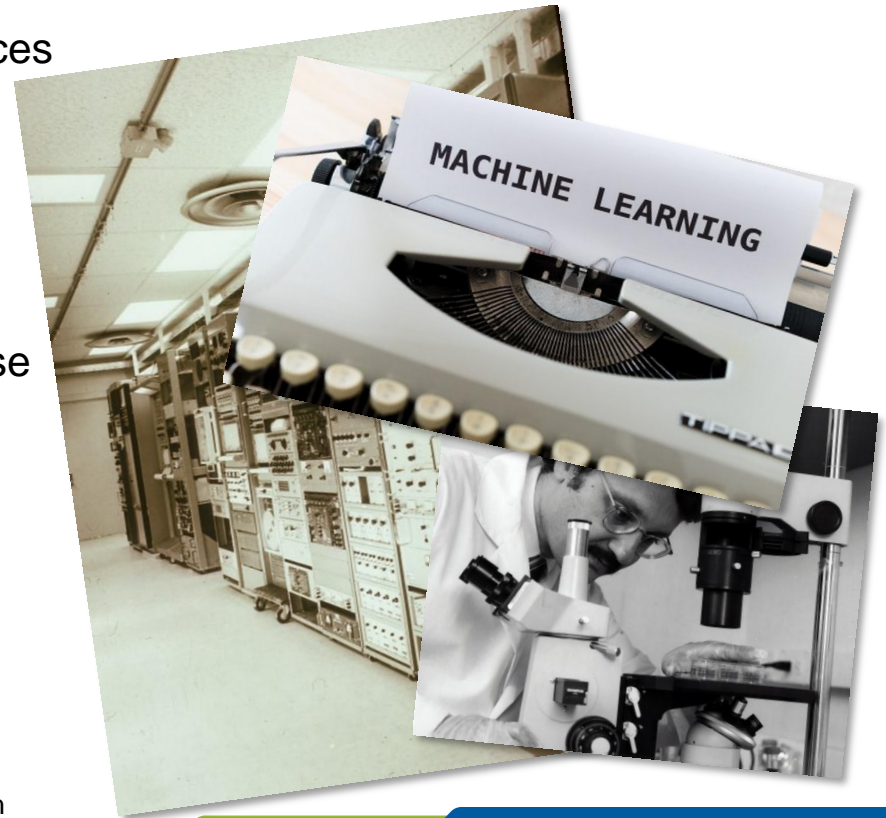
# Why HIFIS?

- Helmholtz aims for joint research & information environment for all Research Fields

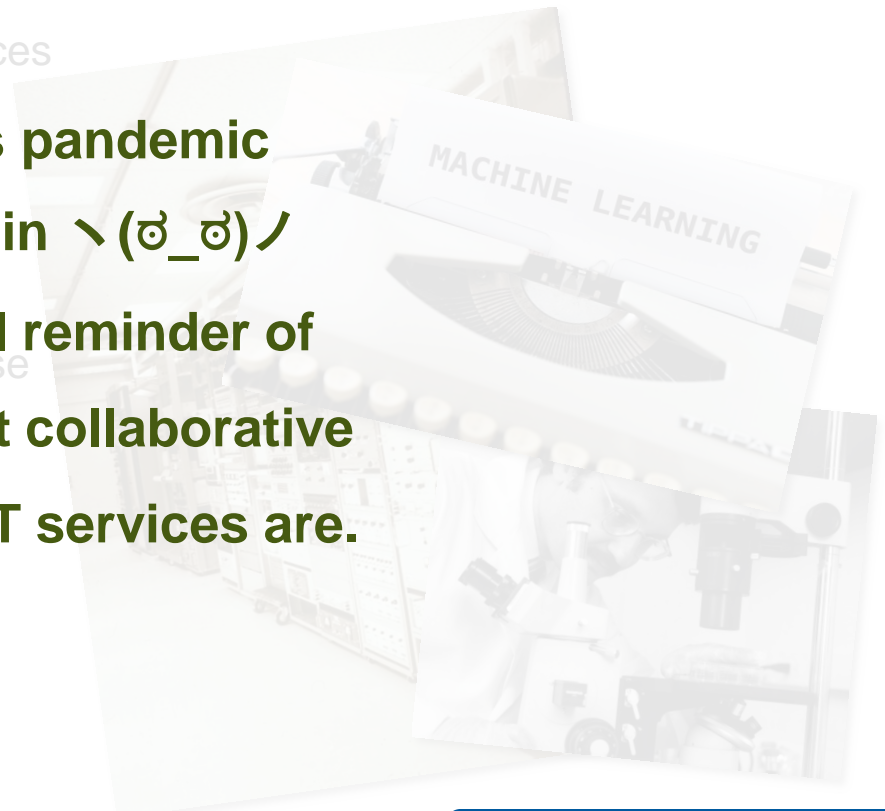


# Why HIFIS?

- Helmholtz aims for **joint research & information environment** for all Research Fields
  - High performance + collaborative services
  - Shall **connect all centres**
    - ....and their world-wide collaboration partners!
  - **Secure, simple access** and easy-to-use
- Widely establish **best-practices** for development + use of research software:
  - high level of quality, visibility and sustainability



- Helmholtz aims for joint research & information environment for all Research Fields
  - High performance + collaborative services
  - Shall **connect all centres**  
...and their world-wide  
collaboration partners!  
**This year's pandemic  
still and again \ (ಠ\_ಠ) /  
is a powerful reminder of  
how important collaborative  
and scalable IT services are.**
  - **Secure, simple access** and easy-to-use
- Widely establish **best-practices** for  
development+use of research software:
  - high level of quality,  
visibility and sustainability



## HIFIS SAB Meeting in April + Report in May

- Open discussions and written feedback report
- High visibility → expert feedback

## Major comments

- Integration with **other platforms + internationally**
- **Integration** into the **proposal process**
- Support scientists + **tackle IT knowledge gap**
- Appreciated HIFIS is a **platform**, not a project

## 18 members participated

from EU (EGI!), USA, AUS:

Christine Kirkpatrick	NDS, SDSC
Ari Asmi	University of Helsinki
Rosa M Badia	BSC
Magchiel Bijsterbosch	SURF
Michael Brüning	University of Queensland
Isabel Campos	CSIC
Tiziana Ferrari	EGI
Andy Götz	ESRF
Christian Grimm	DFN
Marc Heron	DLS
Mirjam van Daalen	PSI
Jörg Herrmann	MPG
Neil Chue Hong	SSI
Rupert Lück	EMBL
Pierre Etienne Macchi	IN2P3 CNRS
Wolfgang E. Nagel	TU Dresden
Davide Salomoni	INFN
Bruno Weikl	FhG

## HIFIS conducted surveys throughout Helmholtz

- Covering all research fields, all centres, scientists/users and ICT provider!
- Software Services survey:

Approximately **1000 replies** from nearly **all centers**

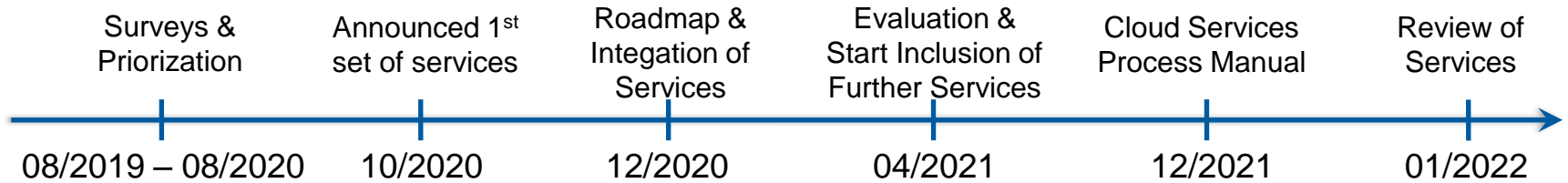
- Cloud Service Survey:

**18+1 centres visited**; Feedback from more than **110 scientists** + ca. 100 IT experts

- Resulted in Initial Service Portfolio







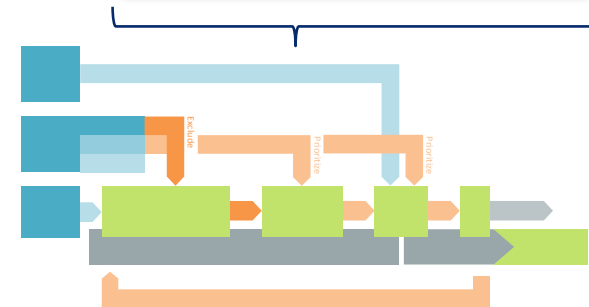
## Done

- **Service Selection** and **Prioritization** finished
  - Process reviewed by Fed Board
  - Strong focus on **benefit for researchers**, and research platforms + International compatibility
  - Full list: [hifis.net → news](https://hifis.net/news)
- Top priority: **Urgent interest** in **collaboration** services and **common access** infrastructure

## Next

- 1<sup>st</sup> “MVP” Version of **Helmholtz Cloud Portal**
- Start / continue **Integration of first services**

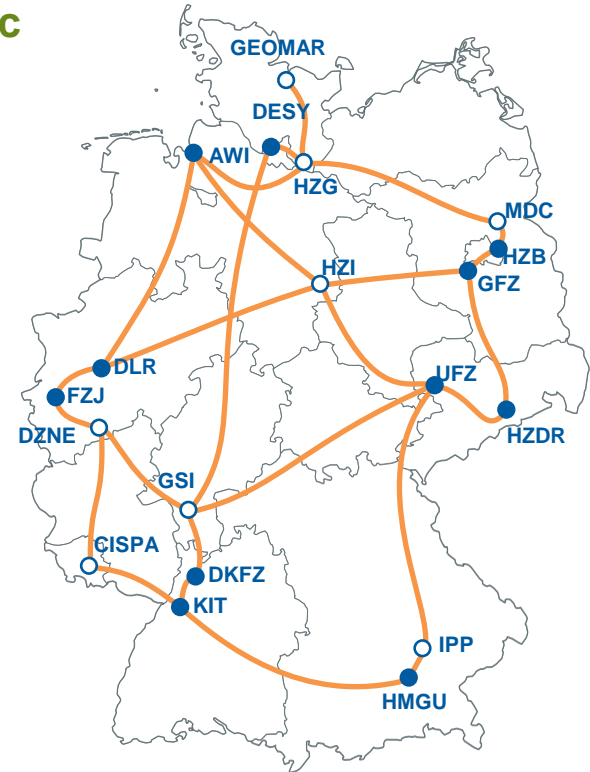
Infrastructure Services		
		indifferent
Storage (HDF)		indifferent'
Storage (HDF)	Storage for HPC,	FZJ
Storage (HDF)	Storage for HPC, general	DESY
Storage (HDF)	Collaboration and Energy	KIT
Communitystorage		indifferent'
Community/ Science Services		
		indifferent
GitHub		Microsoft
GitLab, GitLab-CI		indifferent'
GitLab, GitLab-CI		HZDR
GitLab, GitLab-CI	Software-code online,	WZL

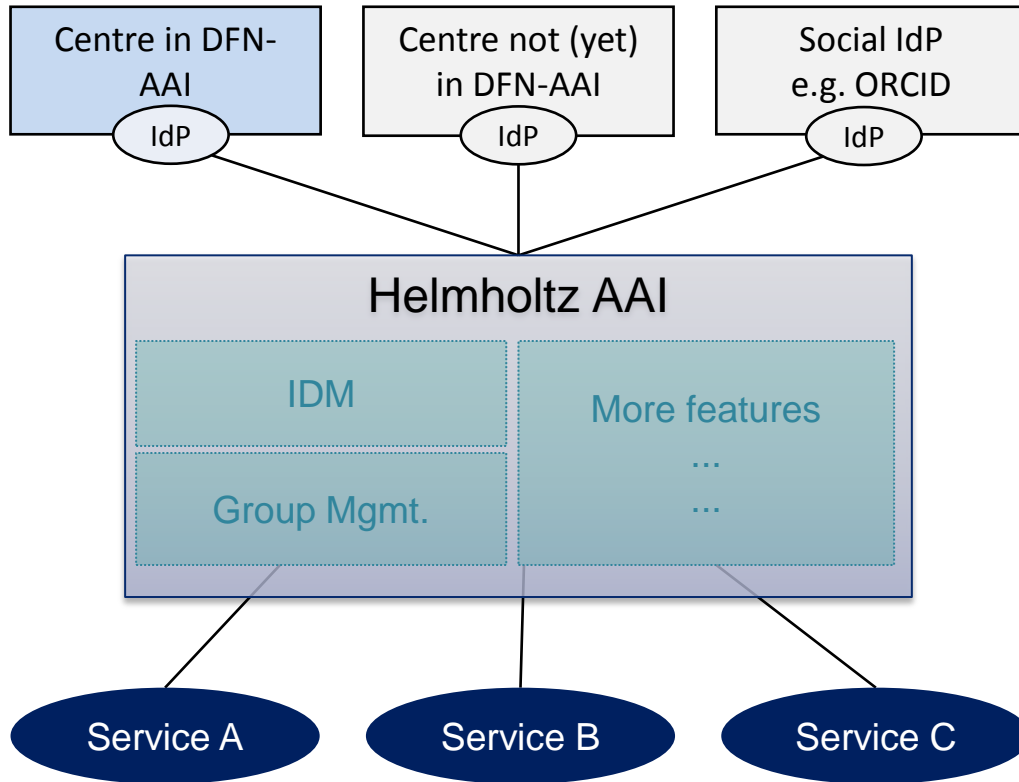




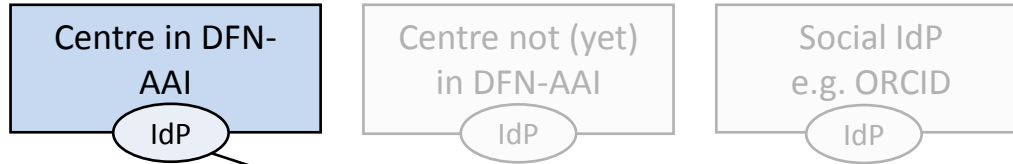
## Common Helmholtz Authentication and Authorization Infrastructure (AAI)

- Enabling **common access to collaborative** and **scientific services** through credentials of the home institution
- **Helmholtz AAI ready** and in use, **compatible to AARC**
- Precondition to build Helmholtz Cloud
  - Various technologies: OIDC, SAML, SSH, LDAP
  - Group / VO Management
  - Secure SSH key management
  - 2FA
  - Deprovisioning (not yet in all modules)
- Documentation and Policies available:
  - <https://hifis.net/doc/backbone-aa1>
  - <https://hifis.net/mission/publications.html>

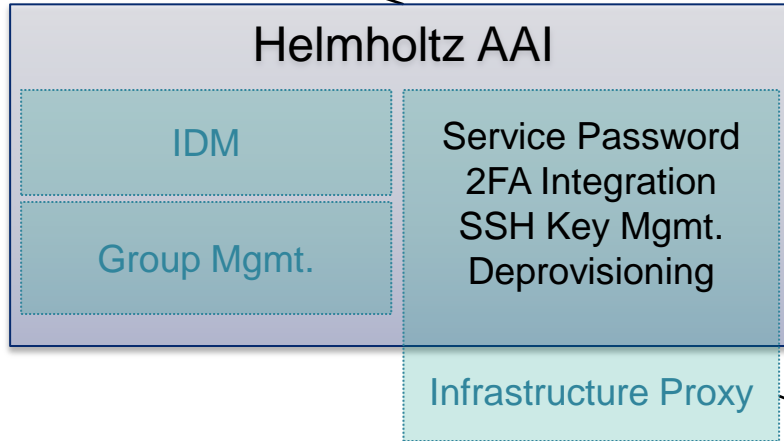




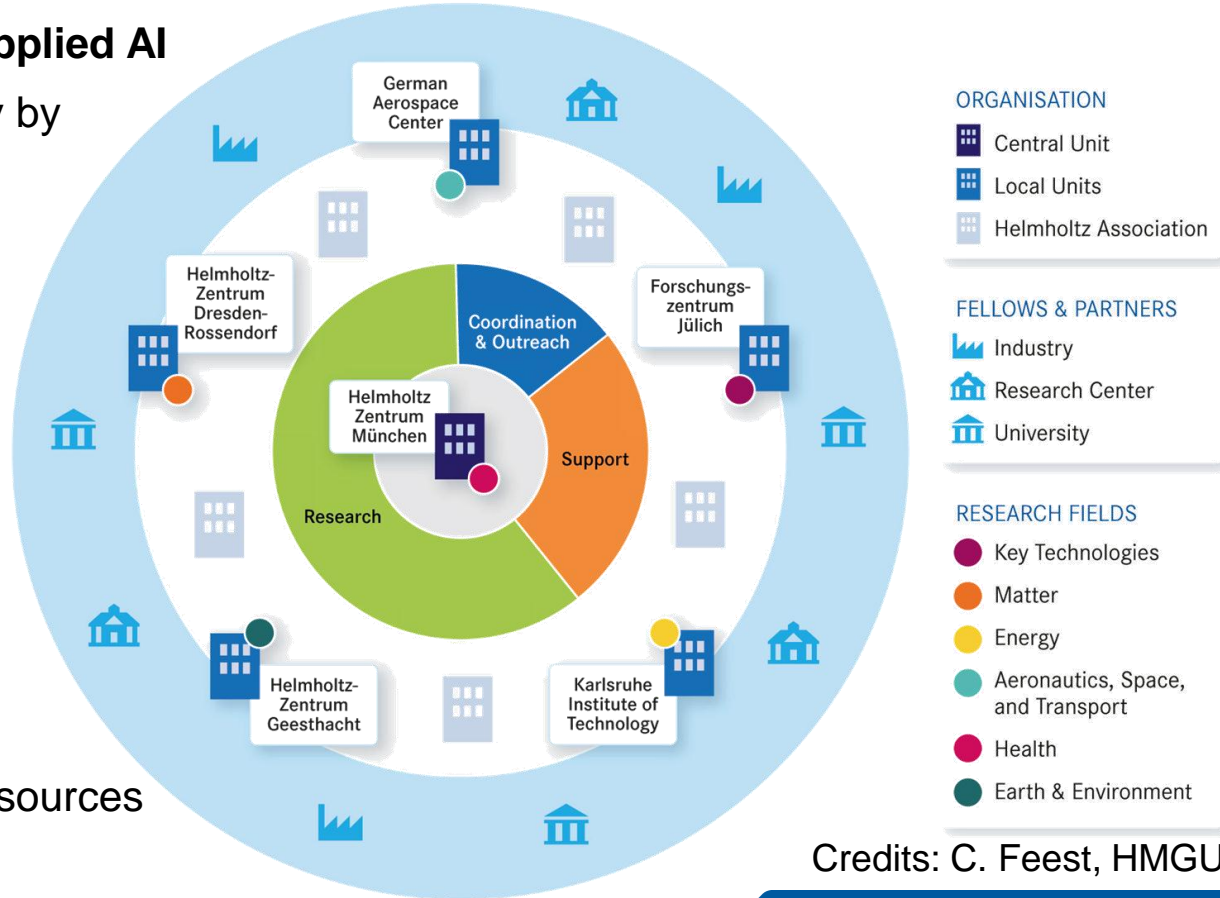
- Helmholtz AAI is a proxy IdP (Identity Provider),
  - following **AARC blueprint**
- Centre IdPs needs to be part of DFN AAI
- Temporary support of IdPs, which are not yet part of DFN AAI
  - Social IdP as fall-back for users without centre IdP
  - User will have lower privileges



- HAICORE Access since July 2020

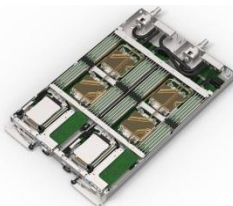


- Research-driven hub for **applied AI**
  - Foster cross-field creativity by **collaborative research projects**
  - Empower & train the current and next generation of scientists
  - Enable efficient & agile development and implementation of AI/ML assets across the whole Helmholtz Association.
- **HAICORE** computation resources
- „Democratizing AI“



Credits: C. Feest, HMGU

- 3x NVIDIA DGX-A100
  - 2x AMD EPYC 64-core CPUs, 1 TB RAM
  - 8x NVIDIA A100 GPUs
  - 40 GB HBM2 memory, 1.5 TB/s
  - 9 Mellanox HDR 200 Gbit/s ports
- 12x Lenovo ThinkSystem SD650Nv2
  - 2x Intel Ice Lake CPU, 512 GB RAM
  - 4x NVIDIA A100 GPUs
  - 1x Mellanox HDR 200 Gbit/s port
- InfiniBand interconnect, 130 GB/s GPFS Storage
- Compute capacity:
  - 72 GPUs x 90% avail. x 24 h x 365 d  
= **~560.000 GPU-h / year**



- Further: J. Buchmüller et al.: Nov 17 10am - 10:30 am  
<https://sc20.supercomputing.org/>

- 2 x 24-core AMD EPYC Rome CPUs
- 4 x Nvidia A100 GPUs, NVLink3
  - 9.7 / 19.5 TF/s peak
  - 40 GB HBM2 memory, 1.5 TB/s
- One Mellanox HDR200 InfiniBand adapter per GPU
- 512 GB DDR4 memory
- DragonFly+ network topology
- 400+ GB/s I/O performance to GPFS-based storage system JUST, up to 1 TB/s to HPST access
- Compute capacity beginning End Nov 2020:
  - 16 nodes x 80% avail. x 4 A100 x 24 h x 365 d  
= **~ 450,000 GPU-h / year**
- Further info: [JuDoor Portal / Jülich](#)

Credits: J. Buchmüller, KIT + N. Attig, FZJ



- 3x NVIDIA DGX-A100
  - 2x AMD EPYC 64-core CPUs, 1 TB RAM

- 2 x 24-core AMD EPYC Rome CPUs
- 4 x Nvidia A100 GPUs, NVLink3

## HAICORE Access board established

- Scientific Approval Process at Helmholtz AI platform
- Involving all Helmholtz stakeholders
- Access policies to be finalized Nov'20

- 12x L
- 2x
- 4x
- 1x

- InfiniE
- Comp

72 GPUs x 90% avail. x 24 h x 365 d  
= ~560.000 GPU-h / year

16 nodes x 80% avail. x 4 A100 x 24 h x 365 d  
= ~ 450,000 GPU-h / year

- Further: J. Buchmüller et al.: Nov 17 10am - 10:30 am  
<https://sc20.supercomputing.org/>

Credits: J. Buchmüller, KIT + N. Attig, FZJ

Projects

Joint application (2 centers)  
'High risk, high gain'

Network

Meetings, Seminars &  
conferences

Research

Young and Senior investigator  
groups

Support

Helmholtz AI consultants  
Voucher system

Training

Hackathons, courses &  
workshops

- Project call #1:
  - [19 groups from all research areas](#)
  - Started early-mid 2020
  - Example →→→→→
- [Project call #2](#):
  - Deadline Dec '20

- HIFIS support for technical implementation of [Voucher system](#)



## ProFile: Protein Folding by Learning

- Launched August 2020 → 2 years

## Aim: predict rigid contact map

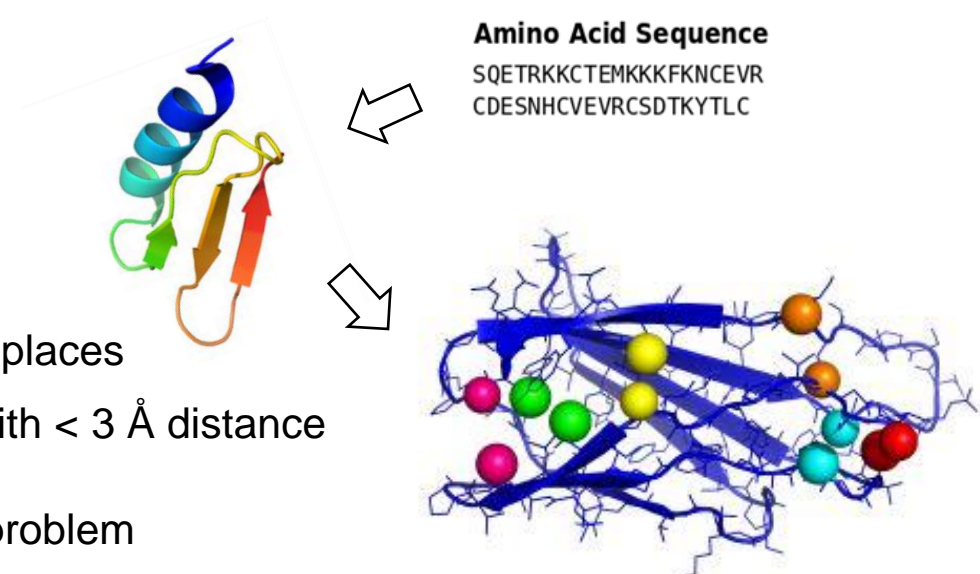
- Input: protein primary structure:  
1D amino acid (AA) sequence with  $n$  places
- Output: 2D contact map – AA pairs with  $< 3 \text{ \AA}$  distance

## Approach: Consider it as a translation problem

- Self-attention neural networks from natural language processing, e.g. Transformer

## Challenge: $O(n^4)$ memory complexity

- Data- and model-parallelization using HPC resources



Credits: M. Götz, KIT

## ProFile: Protein Folding by Learning

- Preprocessing: HTC workload
  - 12 hours job, 100 nodes each 24 cores
  - Roughly 100 GB compressed output
- Sharing **via cloud**:
  - Source code via git repositories at KIT
  - Data via bwSync&Share
- Processing: **HPC** workload on **HAICORE**
  - **A100** NVidia GPUs, NVLink and Infiniband via MPI
  - Custom data-parallel training, **~3x faster than Horovod**
  - **Parallel evolutionary neural architecture search**



# Roadmap (Selection) / Next Steps

- **End Oct'20:** Roadmap for Integration of Initial Cloud Services
- **Nov'20:** Integrate the Software Management Platform into the Helmholtz Cloud
- **Jan'21:** Set-up of Helmholtz Backbone Connections
- **March'21:** Reporting for the first two years of HIFIS: Including User-oriented KPIs
- **April'21:** Start evaluation of further Cloud Services
- **End'21:** Cloud Service Process Manual

➤ <https://hifis.net/roadmap>



**ROADMAP FOR THE INTEGRATION OF THE INITIAL SERVICES** Cloud Services

Having finished the list of the initial service portfolio, the integration of the services will start. As different services need different time for their integration, a specific roadmap will be established.

**November 2020**

**INTEGRATE THE SOFTWARE MANAGEMENT PLATFORM INTO THE HELMHOLTZ CLOUD** Software Services

The software management platform is planned to be integrated into the Helmholtz Cloud as soon as the technical platform is available. Continuous Integration will be part of the software management platform from the very beginning.

**RECIPES FOR A SCALABLE CI SOLUTION**

One important element of the software management platform is a solution for **continuous integration (CI)**. Reusable Ansible recipes for a GitLab CI infrastructure will be provided. The Ansible role is developed [here](#).

**2021**

**January 2021**

**SET-UP OF HELMHOLTZ BACKBONE CONNECTIONS** Backbone Services

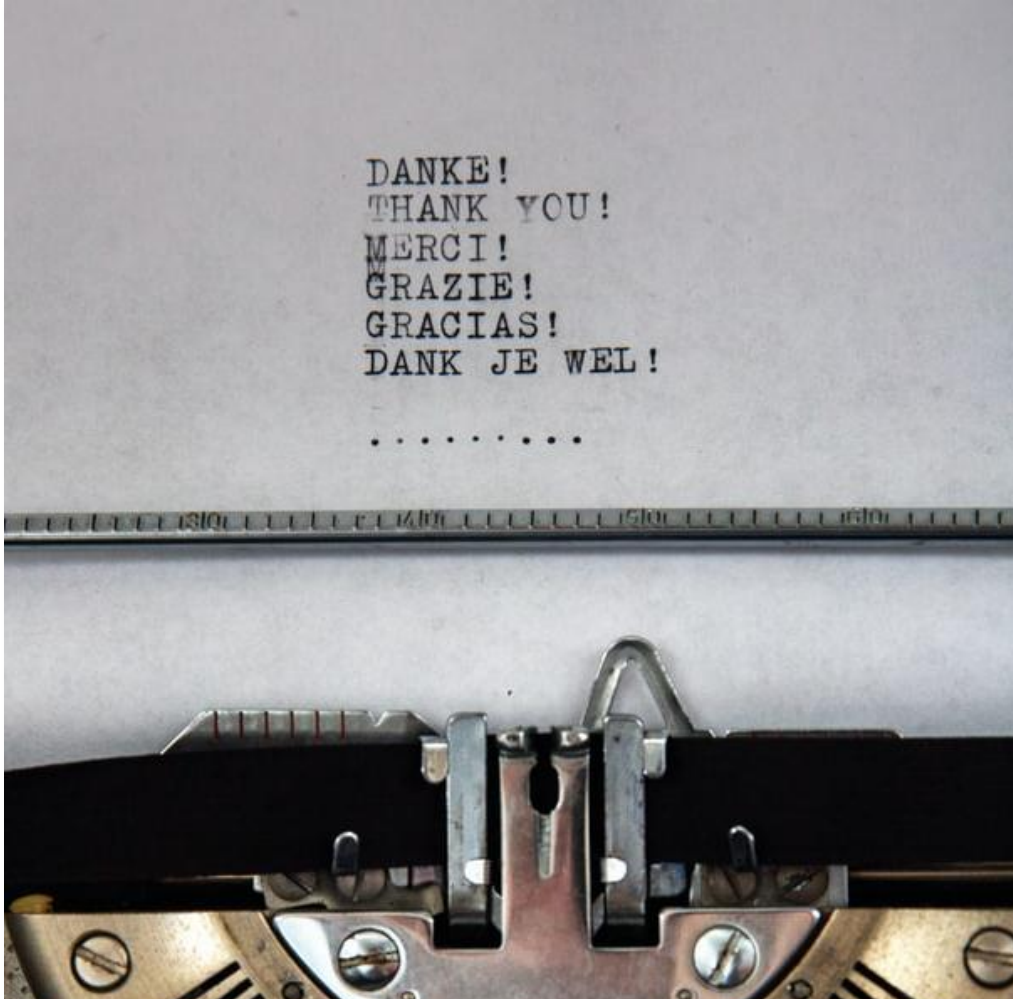
First Virtual Private Network (Helmholtz Backbone) connections have been set up and are in productive use.

## Summary

---



- Very positive and intensive feedback from surveys conducted by HIFIS.
- **Highest interest** in and outside of Helmholtz for **commonly accessible** collaboration tools, high performance IT and sustainable software development.
- Numerous pilot projects throughout 2020 (HZI, DZNE, DLR, HIP, AI, HMC, ...)
- Currently: Integration of initial set of **Helmholtz Cloud IT services**.
- **AAI Connectivity** for all services **and HAICORE**.



DANKE!  
THANK YOU!  
MERC!  
GRAZIE!  
GRACIAS!  
DANK JE WEL!

.....

<https://hifis.net>

- [Subscribe for HIFIS announcement letter!](#)
- General queries to: [support@hifis.net](mailto:support@hifis.net)
- Coordinators: [office@hifis.net](mailto:office@hifis.net)

<https://helmholtz.ai>

- [Subscribe for Helmholtz AI newsletter!](#)