

# GWDG NACHRICHTEN 11-12|23

Machine Learning with HPC

EOSC PID Meta Resolver

Einheitliche  
Filmidentifikatoren

Data Science Summer  
School 2023

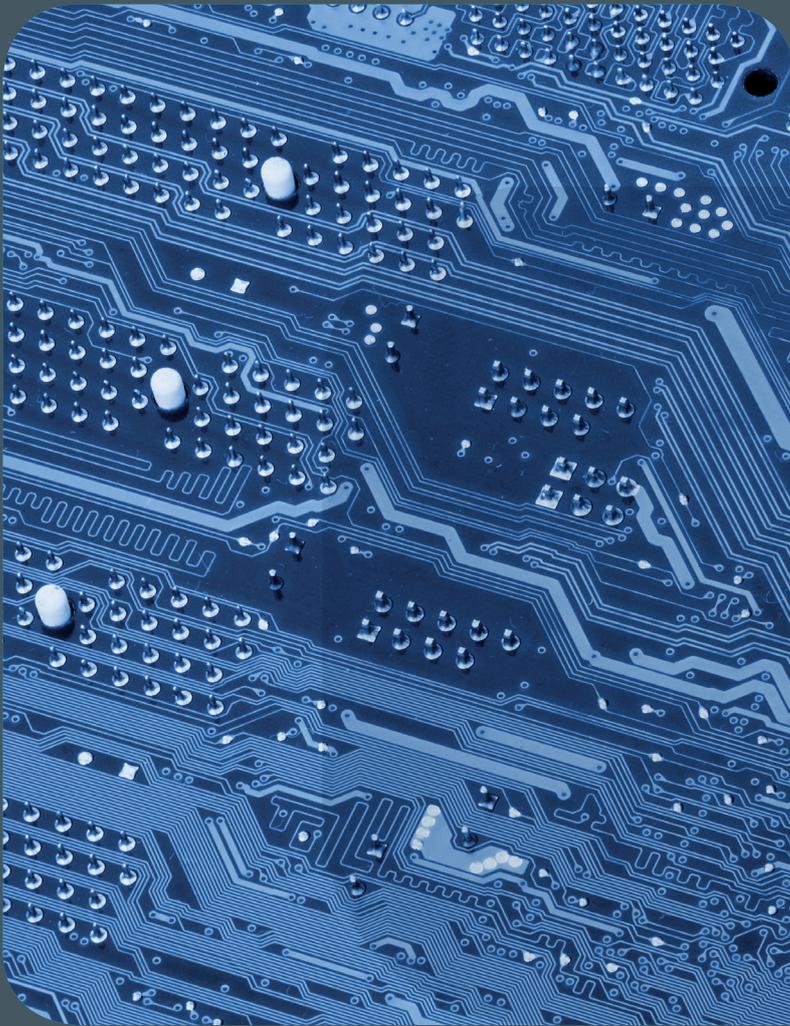
1. NHR-Konferenz 2023

2. IT4Science-Days

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG

***Frohe Weihnachten  
und einen guten  
Rutsch ins neue Jahr!***

 GWDG  
Gesellschaft für wissenschaftliche  
Datenverarbeitung mbH Göttingen



## **GWDG** **NACHRICHTEN**

# 11-12|23 Inhalt

- .....
- 4 **Best Practices for Machine Learning with HPC**
  - 8 **FAIRCORE4EOSC – EOSC PID Meta Resolver**
  - 12 **Projektstart zu einheitlichen Filmidentifikatoren in einem Verbundsystem** 16 **Start der „Landesinitiative Forschungsdatenmanagement Niedersachsen“** 18 **Data Science Summer School 2023** 21 **Nationales Hochleistungsrechnen – Vorhang auf zur 1. NHR-Konferenz 2023** 23 **„Bei Euch wird es auch immer schwieriger?“ – Die 2. IT4Science-Days der Helmholtz-IT und der Max-Planck-IT**
  - 25 **Kurz & knapp** 27 **Stellenangebot**
  - 28 **Personalia** 29 **Academy**

### Impressum

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686  
46. Jahrgang  
Ausgabe 11-12/2023

**Erscheinungsweise:**  
10 Ausgaben pro Jahr

[www.gwdg.de/gwdg-nr](http://www.gwdg.de/gwdg-nr)

**Auflage:**  
550

**Fotos:**  
© sosiukin - stock.adobe.com (1)  
© intheskies - Fotolia.com (11)  
© wladimir18 - stock.adobe.com (14)  
© chagin - Fotolia.com (15)  
© Sarah Diemert (17)  
© Nataliya Kalabrina - Fotolia.com (20)  
© ALOI.PHOTO (21, 22)  
© Patrick Schmidt (24)  
© pineapple - Fotolia.com (26)  
© nito - Fotolia.com (27)  
© Robert Kneschke - Fotolia.com (29)  
© MPI-NAT-Medienservice (3)  
© GWDG (2, 18, 19, 28)

**Herausgeber:**  
Gesellschaft für wissenschaftliche  
Datenverarbeitung mbH Göttingen  
Burckhardtweg 4  
37077 Göttingen  
Tel.: 0551 39-30001  
Fax: 0551 39-130-30001

**Redaktion und Herstellung:**  
Dr. Thomas Otto  
E-Mail: [thomas.otto@gwdg.de](mailto:thomas.otto@gwdg.de)

**Druck:**  
Kreationszeit GmbH, Rosdorf



Prof. Dr. Ramin Yahyapour  
ramin.yahyapour@gwdg.de  
0551 39-30130

## *Liebe Kund\*innen und Freund\*innen der GWDG,*

*mit dem Jahr 2023 kann man sehr unterschiedliche Ereignisse und Entwicklungen verbinden. Auf der einen Seite sind wir global weiterhin mit Polykrisen konfrontiert, deren Ende und langfristigen Auswirkungen nicht absehbar sind. Auf der anderen Seite haben sich Herausforderungen aus 2022 zu Energieengpässen und Lieferproblemen entspannt. Für die GWDG war sicherlich der Sicherheitsvorfall aus dem Herbst ein bemerkenswertes Ereignis. Auch wenn wir diesen vergleichsweise gut überstanden haben, sind wir noch dabei, die Lehren und Maßnahmen daraus abzuleiten. Der Ausbau unseres Sicherheitsmanagements wird uns in 2024 sicherlich beschäftigen. Dazu passt, dass wir gerade unser Überwachungsaudit für die zugehörige ISO-27001-Zertifizierung sehr gut durchlaufen haben. Zum Abschluss dieses Jahres kam noch die positive Meldung, dass die GWDG bei einer Ausschreibung der European Open Science Cloud als Dienstleister ausgewählt wurde.*

*So gehe ich auch für 2024 von neuen Herausforderungen aus, die wir in bewährter Art und Weise meistern werden. Ich wünsche Ihnen und Ihren Familien schöne Feiertage und einen erfolgreichen Start in das Jahr 2024.*

**Ramin Yahyapour**

*GWDG – IT in der Wissenschaft*

# Best Practices for Machine Learning with HPC

**Text and Contact:**

Ali Doost Hosseini  
ali.doost-hosseini@gwdg.de  
Hauke Kirchner  
hauke.kirchner@gwdg.de  
Dorothea Sommer  
dorothea.sommer@gwdg.de

The GWDG offers data scientists various services and training courses to support them in their work throughout their entire workflow. As the success of a Machine Learning project often depends on the available computing resources, the working group “Computing” operates various HPC systems with suitable accelerators to train Deep Learning models. This article explains model training using the highly energy-efficient HPC cluster “Grete”. The example is an excerpt from the course “Deep Learning with GPU cores,” offered every six months.

## INTRODUCTION

The GWDG offers several services and trainings to support researchers in their machine-learning projects [1]. Deep Learning success is heavily compute-bound [2], [3]; accelerators integrated into our HPC systems can help overcome this bottleneck. The GWDG offers several HPC systems [4] with modern GPU nodes and other accelerators [5] accessible to different user groups [6]. If you are wondering whether you have access to our systems, look at our science domains blog article from June [6]. Additionally, the KISSKI project [7] is not limited to researchers, but small and medium-sized enterprises are highly welcome.

Further, we aim to support different science domains in the scope of the NHR alliance [8]. Here we focus on researchers in the fields of artificial intelligence [9], digital humanities, bioinformatics [10], and forest science [11].

## MACHINE LEARNING WORKFLOW

A typical machine-learning workflow can be seen in Figure 1. However, this article will only focus on model development and evaluation of best practices. Regardless, the GWDG offers services to support you in every step of this workflow [1]. If you need

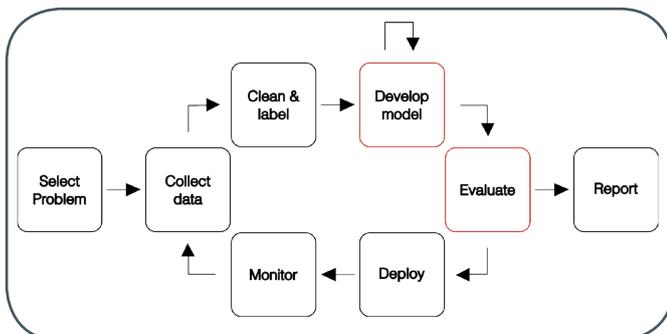


Figure 1: A typical machine learning workflow involves data collection, pre-processing, dataset building, model training and refinement, evaluation, deployment, and continuous monitoring and updating. This process is iterative and may require revisiting previous steps based on the results at later stages [12], [13], [14], [15], explained by a transformer model deployed by Perplexity [16].

our assistance with anything, feel free to contact us.

## HPC Cluster

A high-performance computing (HPC) cluster consists of different essential components, which are critical to understand to use its resources efficiently. Here, we will briefly explain the basic terminology. If you are interested in a more detailed article on this topic, please read our science domain blog article “What scientists should know to efficiently use the Scientific Computing Cluster” [17].

First, a user can access the login node of the HPC cluster via an SSH (Secure Shell Protocol) connection (1 in Figure 2).

In the example workflow, you can see that *glogin9* is the login node of our GPU system Grete [18]. On the login node, jobs are submitted with the job scheduler SLURM [19] (2 in Figure 2). A job on SLURM is a combination of which program you would like to run and which computing resources (hardware) you need.

Compute nodes take care of the actual computing of your jobs, similar to your personal computer. Those compute nodes consist of cores and memory (RAM). Cores are processing units, such as CPU or GPU. Each node can have a different number of cores. Each node also has its own (temporary) random-access memory (RAM) used for temporary computations. Various storage

## Bewährte Praktiken für maschinelles Lernen mit HPC

Datenwissenschaftler\*innen bietet die GWDG diverse Dienstleistungen und Trainings an, um sie bei ihrer Arbeit über den gesamten Arbeitsprozess hinweg zu unterstützen. Da der Erfolg eines Projektes zum maschinellen Lernen oft von der verfügbaren Rechenleistung abhängt, betreibt die Arbeitsgruppe „Computing“ (AG C) verschiedene HPC-Systeme mit geeigneten Beschleunigern für das Training von Deep-Learning-Modellen. Am Beispiel des sehr energieeffizienten HPC-Clusters „Grete“ wird das Modeltraining konzeptuell erklärt. Dies ist ein Auszug aus dem Kurs „Deep Learning with GPU cores“, der halbjährlich angeboten wird.

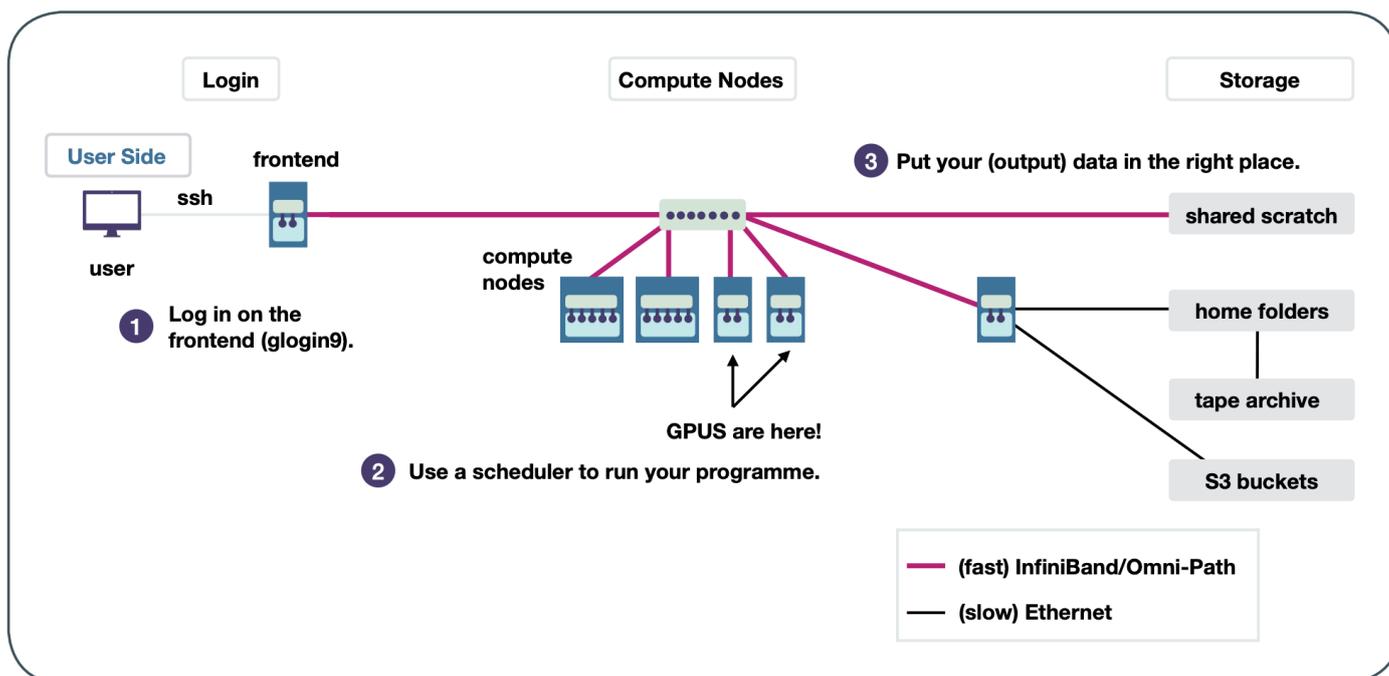


Figure 2: Schematic visualization of an HPC system, similar to the system “Grete”. On the left is the user and the front nodes that are used for login. In the center, the compute nodes handle the computation of the submitted jobs. Those are connected with different storage systems on the right. The Figure was taken from our “Deep Learning with GPU” course offered every half year [20].

systems with different characteristics help to handle and organize data efficiently (3 in Figure 2).

While using the *shared scratch* for your computation has great results, it is not great for long-term storage. This type of file system has no backup! However, your *home* folder has a backup. As indicated in the Figure by the connection to the tape archive (right center). These are literal tapes that, after being written, are stored as physical components not connected to any electricity. Lastly, connecting S3 buckets to the whole cluster (bottom right) is also possible.

## CODING

In order to use the HPC resources, you will need to write code and develop programs. The programming language depends on the target application. For machine learning and deep learning, Python is often the best choice due to its moderate learning curve and availability of comprehensive packages and frameworks, such as PyTorch, JAX (with Flux) or Tensorflow.

### Tools and Frameworks

For development, we recommend VSCode [21] as it supports many features and programming languages including Python through its vast selection of extensions [22]. VSCode can also be configured to connect to the cluster directly which makes development simple and straightforward [23].

When working on a Python project, it is best to use a virtual environment or container with a tool such as conda, singularity or apptainer [24], to ensure reproducibility and robustness of the code, as packages are constantly updated and your code may require specific versions.

There are several widely-used frameworks for Deep Learning in Python. Three more common ones are PyTorch, TensorFlow, and JAX. Tensorflow is the first major Deep Learning framework and was developed by Google in 2015. PyTorch was developed in

2016 by Facebook, and is currently the most dominant framework in academia winning 77% of the competitions. JAX [25] is a relatively new framework developed by Google that is aimed towards a more general auto-differentiation and vectorisation framework. For most cases, we recommend using PyTorch.

### Documentation and Code Handling

Code should be well-documented and clean; comments and README files should be included when necessary and repetitive code must be converted into functions and packages. It is also recommended to treat raw data as immutable; build the code around the data instead of editing the raw data itself. Cookiecutter is a tool that can get you started with a decent project structure and has some good recommendations as well [26].

It is crucial to use a Version Control System (VCS) such as git to avoid accidents and disasters when altering code and in general to gain better access and control over the code history. GWDG offers a GitLab server for this purpose. A good example is the “Deep Learning with GPU” workshop repository located in GWDG’s own GitLab server [20].

### Job Scheduling with SLURM

Running complex programs is forbidden on the login nodes. It is important to run complex tasks, e.g. data processing, model training and testing, only on the compute nodes. As mentioned before, the login nodes are strictly for low-cost tasks and job submission. You must use our job scheduler SLURM to run jobs on the various nodes in the cluster. When submitting a job via SLURM, you must set the job configuration, e.g. requested partition/node and GPUs, and you may also add additional settings such as the amount of RAM, time limit, email address to be notified when the job is complete, and a location to store the job’s log file produced by slurm. It is recommended to specify a location to store the log file and regularly check it for any issues. An overview of the configurations in SLURM can be found in our documentation [27].

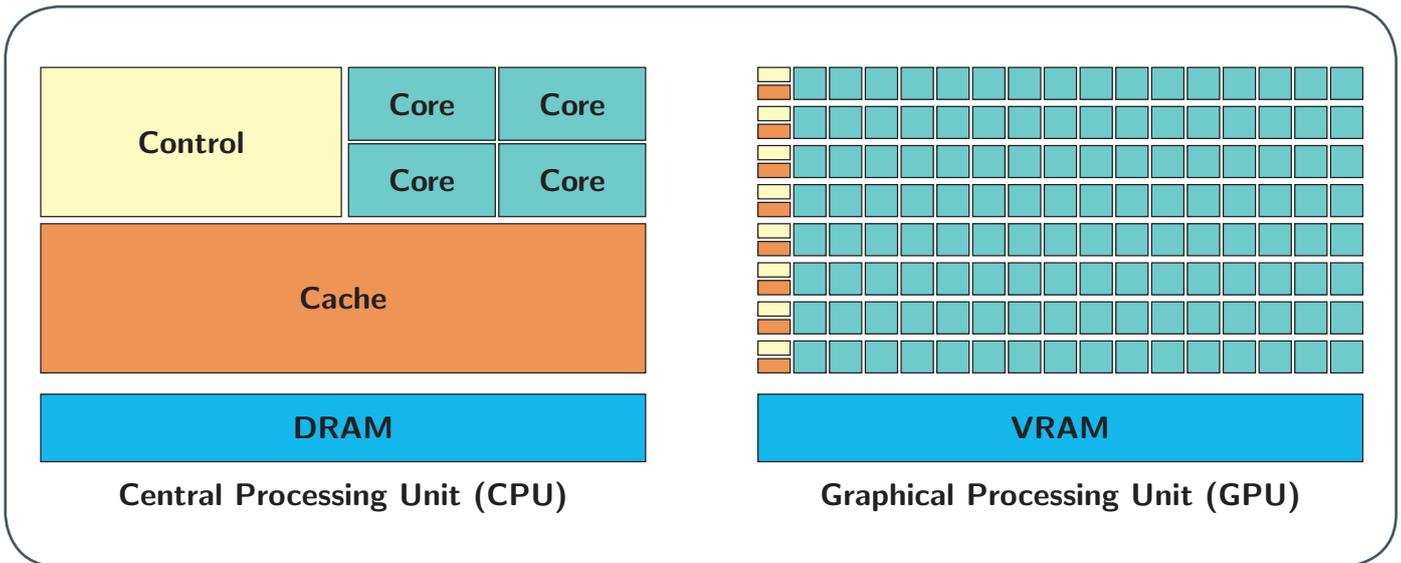


Figure 3: Internal comparison between a CPU and a GPU

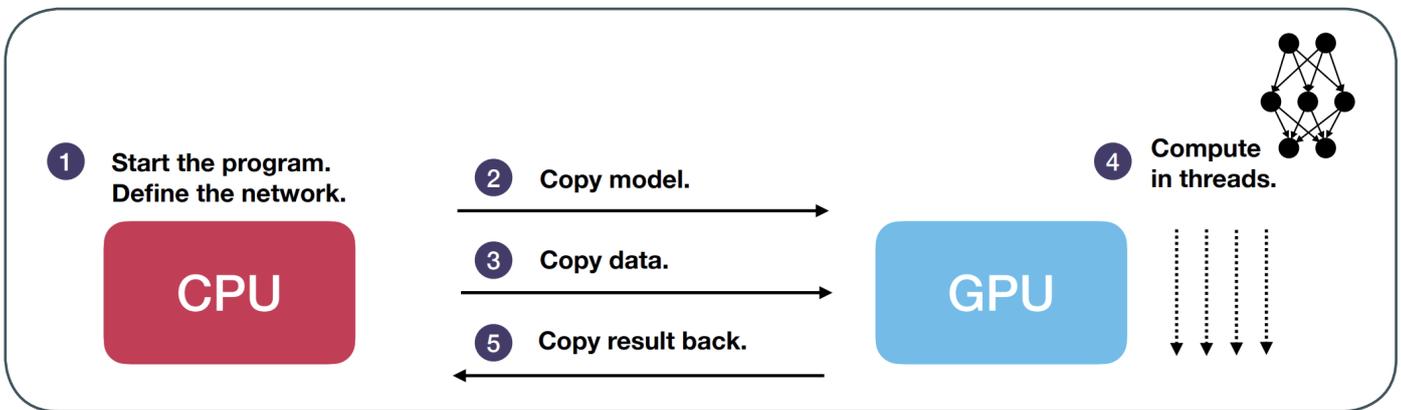


Figure 4: Role of CPU and GPU in typical deep learning task

During model development, you may want to run interactive jobs in the terminal with *srn*, e.g., for testing purposes. In this case it is better to use nodes in partitions designated for this use case such as *grete-interactive*. Finally, you may train the model by writing a SLURM batch script and running it with *sbatch* on a dedicated GPU node, e.g. on partition *grete*. [28]

You must also make sure that the requested node configuration is available and suitable for the job. For example, during development if the job is more likely to fail, it is better to set a low time limit and run the job on a test node if applicable. Similarly, the number of nodes and resources must be sufficient for your job while not being too excessive. It is better to split a long job into smaller jobs to avoid time limit issues and potentially wasting resources. To chain consecutive jobs with automated start and easier check-pointing we recommend *snakemake* [29].

### MONITORING

While your program is running, it is important to monitor the resource utilization on the nodes to identify bottlenecks, errors, and possible improvements.

### CPU vs. GPU

CPUs and GPUs are both powerful processing hardware that can perform complex calculations. Figure 3 depicts the inner

structure of CPUs and GPUs. While CPUs are capable of handling a larger variety of instructions, GPUs are more efficient when it comes to specific simple instructions. It is important to know when using a GPU improves performance and when it does not.

A thread is a series of sequential operations. A CPU is capable of processing multiple threads performing different tasks simultaneously. GPUs, on the other hand, are generally suitable for tasks that run the same operation on a large number of elements, i.e., running many similar threads at the same time.

Figure 4 depicts a typical Deep Learning task running on a GPU node. The CPU is responsible for starting and running the program, beginning by initializing the model and storing it in the RAM. We would now like to take advantage of the GPU's fast processing power in order to train the model, however, the GPU cannot directly access data stored in the RAM. Instead, it has its own memory (usually called VRAM, short for Video-RAM), which can vary in size depending on the GPU, but is typically smaller than the RAM.

During training, the model parameters and data must be copied from the RAM into the GPU's VRAM. The GPU can then feed the input data into the model, compute the result, and update the model. The result may then be transferred back to the CPU for further processing.

Transferring data between the CPU and GPU is not immediate, and the time it takes should be taken into account. If the processing task is too simple, running it on a GPU would have little

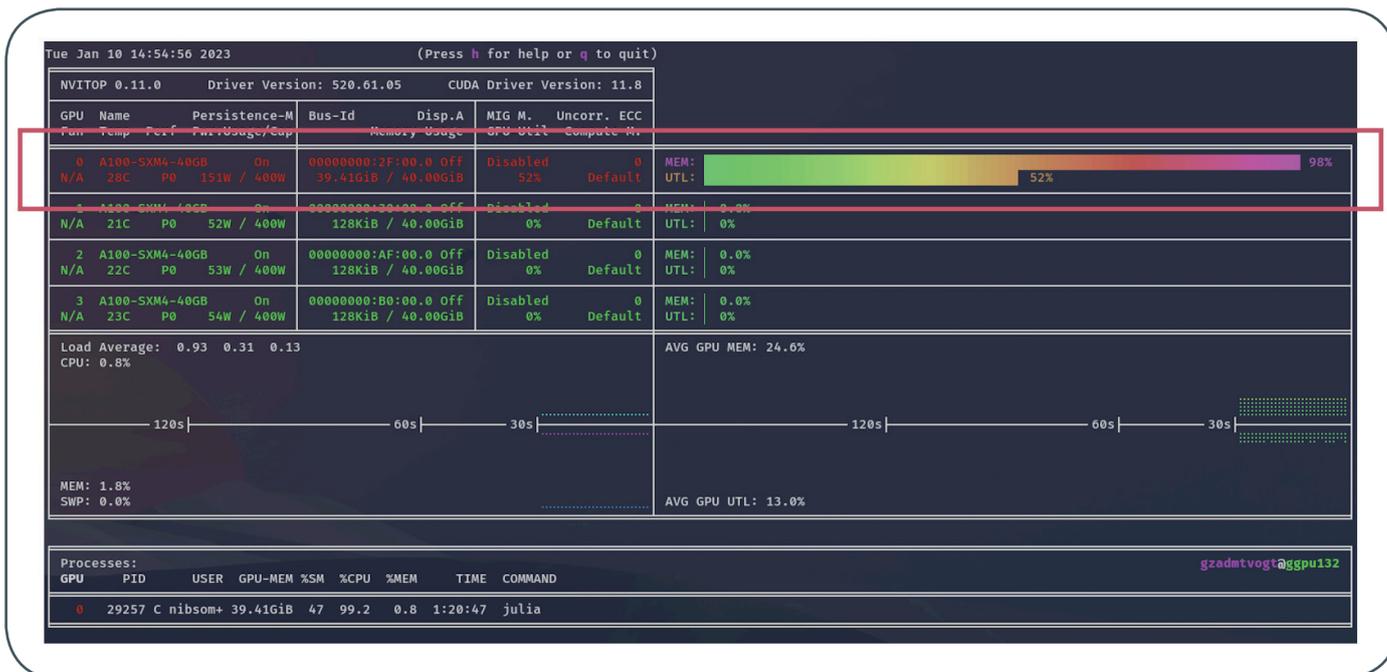


Figure 5: Example output from `nvidia-smi`

advantage over a CPU. To make full use of a GPU, you must ensure that lots of computation is performed on the GPU and the amount of data transfer is low in comparison. This can be verified using our recommended monitoring tools.

### CPU and RAM Usage

It is possible to view the hardware resource usage of a node by simply logging into it and running `top` (for CPU nodes) or `module load nvidia-smi` then `nvidia-smi` (for GPU nodes) in the terminal.

A typical output of `nvidia-smi` is shown in Figure 5. MEM depicts the amount of allocated GPU memory and UTL shows the GPU utilization. A constant low utilization of the GPU means that the GPU is mostly idle and waiting for more data.

The GPU memory should be large enough to support the model and batch of data simultaneously. There are different types of GPUs available on the cluster which can be selected depending on the use case. For newer GPUs, you may run your job on a GPU slice using NVIDIA-MIG when the full GPU power is not required. An overview of the available GPUs and usage can be found under [28].

### Example

An example of Deep Learning with GPU on the cluster and a recorded video series is available [20], [30].

### LINKS AND FOOTNOTES

[1] <https://gwdg.de/en/services/>  
 [2] <https://openai.com/research/ai-and-compute#addendum>  
 [3] <https://academy.gwdg.de/academy>  
 [4] <https://gwdg.de/hpc/systems/>  
 [5] Both graphcore systems and neuromorphic chips are being acquired as part of the KISSKI project  
 [6] [https://gitlab-ce.gwdg.de/hpc-team-public/science-domains-blog/-/blob/main/20230623\\_hpc-access.md?ref\\_type=heads](https://gitlab-ce.gwdg.de/hpc-team-public/science-domains-blog/-/blob/main/20230623_hpc-access.md?ref_type=heads)

[7] <https://kisski.gwdg.de/>  
 [8] <https://www.nhr-verein.de/en>  
 [9] <https://gwdg.de/en/community-pages/ai-intro/>  
 [10] <https://gwdg.de/en/community-pages/bioinfo-intro/>  
 [11] <https://gwdg.de/en/community-pages/forestry-intro/>  
 [12] <https://www.run.ai/guides/machine-learning-engineering/machine-learning-workflow>  
 [13] <https://ml-ops.org/content/end-to-end-ml-workflow>  
 [14] <https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94>  
 [15] <https://cloud.google.com/ai-platform/docs/ml-solutions-overview>  
 [16] <https://www.perplexity.ai/>  
 [17] [https://gitlab-ce.gwdg.de/hpc-team-public/science-domains-blog/-/blob/main/20230417\\_cluster-practical.md](https://gitlab-ce.gwdg.de/hpc-team-public/science-domains-blog/-/blob/main/20230417_cluster-practical.md)  
 [18] <https://www.hlm.de/doc/display/PUB/GPU+Usage>  
 [19] <https://slurm.schedmd.com>  
 [20] <https://gitlab-ce.gwdg.de/dmuelle3/deep-learning-with-gpu-cores>  
 [21] <https://code.visualstudio.com/>  
 [22] <https://marketplace.visualstudio.com/VSCode>  
 [23] <https://info.gwdg.de/news/en/configuring-vscode-to-access-gwdgs-hpc-cluster/>  
 [24] [https://gitlab-ce.gwdg.de/hpc-team-public/science-domains-blog/-/blob/main/20230907\\_python-apptainer.md?ref\\_type=heads](https://gitlab-ce.gwdg.de/hpc-team-public/science-domains-blog/-/blob/main/20230907_python-apptainer.md?ref_type=heads)  
 [25] <https://jax.readthedocs.io/en/latest/index.html>  
 [26] <https://drivendata.github.io/cookiecutter-data-science/>  
 [27] <https://www.hlm.de/doc/display/PUB/Slurm>  
 [28] <https://www.hlm.de/doc/display/PUB/GPU+Usage>  
 [29] <https://snakemake.github.io/>  
 [30] <https://www.youtube.com/playlist?list=PLvcoSsXFNRbIM4AG5PZwY1AfYEW3EbD90>

# FAIRCORE4EOSC – EOSC PID Meta Resolver

## Text and Contact:

Ali Reza Sajedi  
ali.sajedi@gwdg.de

Dr. Sven Bingert  
sven.bingert@gwdg.de

The PID Meta Resolver (PIDMR) is a generalized resolver for mapping items into records which is to be developed as one of the nine components defined within the framework of the FAIRCORE4EOSC project [1]. The PIDMR will know where to route different types of Persistent Identifier (e.g. DOI or URN:NBN). Usage of the PIDMR should improve machine based data processing and allow retrieving digital object information without in-depth knowledge of the resolution mechanism of different PID systems.

## THE PERSISTENT IDENTIFIER META RESOLVER – PIDMR

The increasing adoption of Persistent Identifiers (PIDs) to reference various types of research findings represents a significant advancement in addressing future demands for enhancing the FAIR (Findable, Accessible, Interoperable, and Reusable) aspects of research data. While this trend is highly commendable, it also brings forth new challenges related to managing and incorporating these PIDs into diverse research workflows. The primary hurdle lies in the diverse array of systems responsible for generating and maintaining these PIDs. The key challenge is determining which system is accountable for the resolution process, which, in turn, furnishes the associated (meta)data for a given PID. The introduction of a standardized interface capable of resolving PIDs from different systems can greatly streamline various processes. The Persistent Identifier Meta Resolver (PIDMR) serves as such an interface and will be expanded within the FAIRCORE4EOSC project to facilitate the integration of various systems.

## ARCHITECTURAL DESIGN

The design (see Figures 1 and 2) follows the methodology of the C4 Model [2].

The PIDMR serves as a versatile resolver designed to associate items with records. It possesses the capability to intelligently direct various types of identifiers, such as DOIs or URNs (e.g., URN:NBN), to their respective destinations. This innovative resolver enhances automated data processing, enabling access to digital object information without necessitating an intricate understanding of the resolution processes employed by diverse PID systems. Consequently, it facilitates the aggregation and analysis of data collections, which may originate from various sources and reference different PID systems.

The PIDMR comprises the following three elements:

1. Backend: This primary component is tasked with resolving a PID and displaying the associated metadata.
2. API Services: REST APIs supporting the core functionalities of the Meta Resolver, facilitating machine

actionability.

3. User Interface (UI): A user-friendly and intuitive interface designed to accommodate various user types for Meta Resolver utilization and retrieving PID service provider information.

## FEATURES

- Centralized resolution of PIDs with associated metadata retrieval for the referenced resource (if supported by the PID System)
- Implementation of a versatile and scalable meta-resolver capable of accessing digital objects referenced by various PID systems
- Establishment of a generic and sustainable infrastructure for managing information about digital objects
- Integration of support for a curated list of widely recognized PID systems
- User-friendly interface with consistent functionality for all users
- Provision of API endpoints to enable integration with the PIDMR by other systems

## EOSC PID Meta Resolver

Der PID Meta Resolver (PIDMR) ist ein verallgemeinerter Resolver für die Zuordnung von Objekten zu Datensätzen, der als eine der neun im Rahmen des FAIRCORE4EOSC-Projekts [1] definierten Komponenten entwickelt werden soll. Der PIDMR weiß, wohin verschiedene Arten von persistenten Identifikatoren (z. B. DOI oder URN:NBN) weitergeleitet werden müssen. Der Einsatz des PIDMR soll die maschinelle Datenverarbeitung verbessern und das Abrufen digitaler Objektinformationen ohne tiefgreifende Kenntnisse der Auflösungsmechanismen verschiedener PID-Systeme ermöglichen.

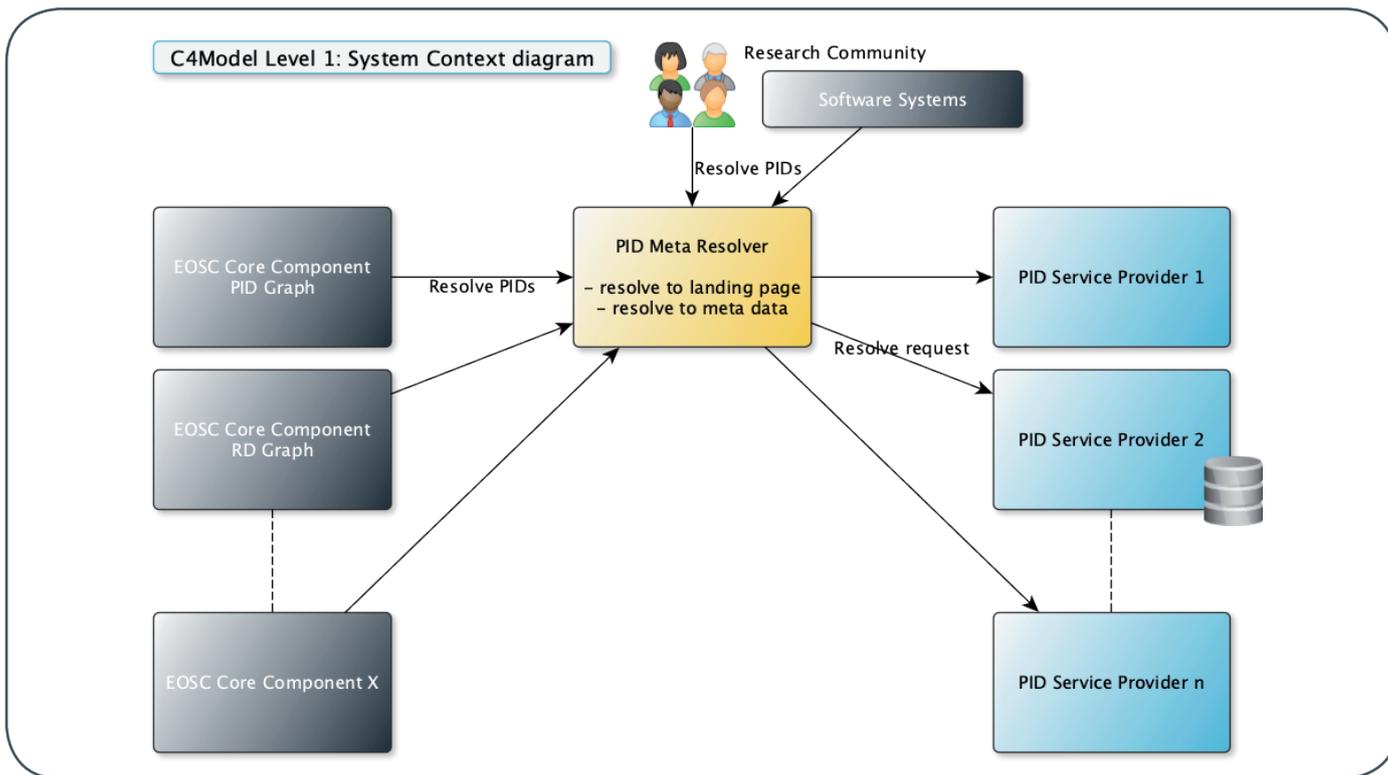


Figure 1: C4 Model Level 1 – System Context Diagram

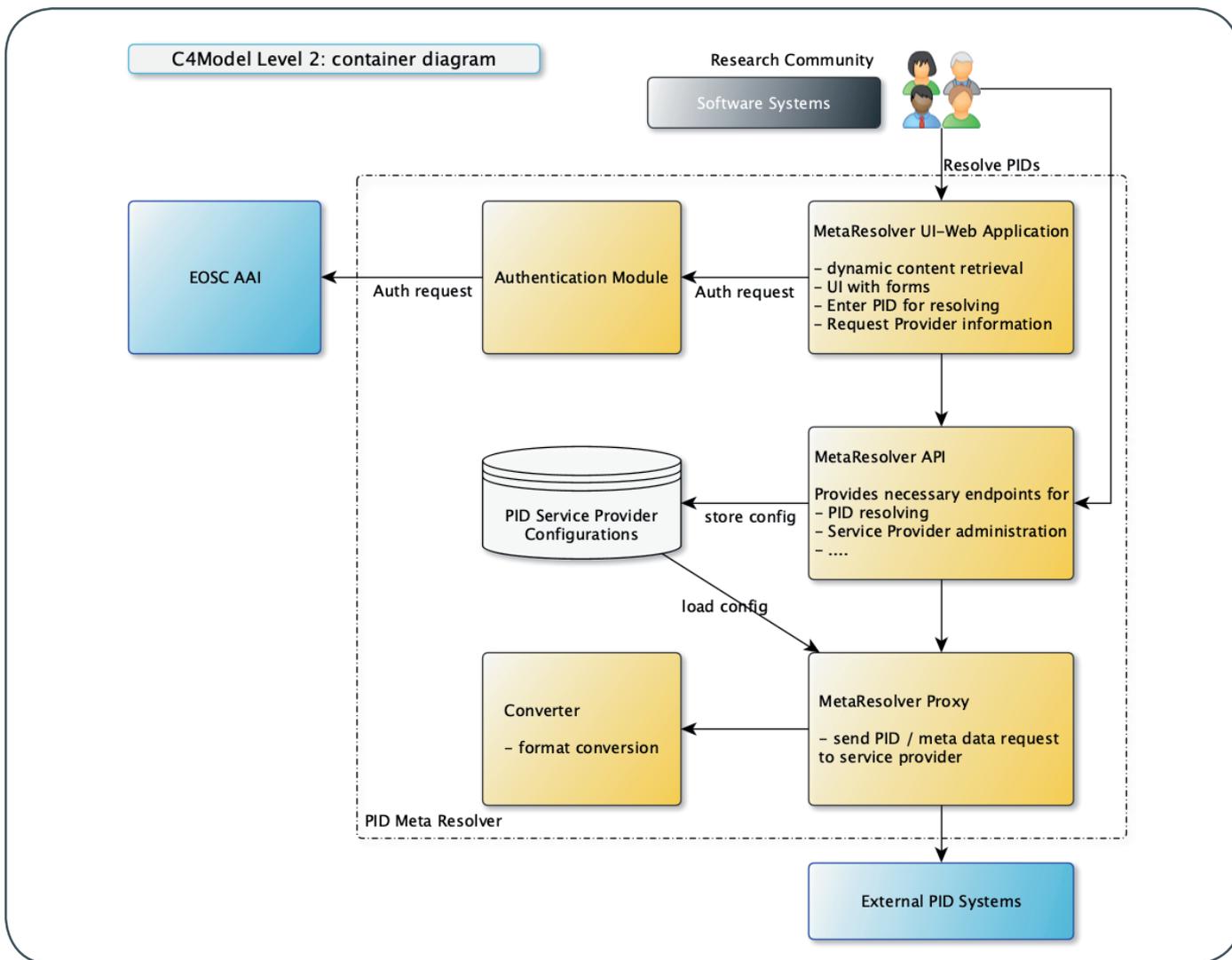


Figure 2: C4 Model Level 2 – Container Diagram

## IMPACT

Facilitate researchers in their everyday tasks by providing seamless access to PIDs (resolution, metadata). Offer a centralized resolution service for PIDs from diverse providers to streamline workflows effortlessly.

## META RESOLVER BACKEND

As part of the milestones of the FAIRCORE4EOSC project, a beta version of all components was carried out at the end of November 2023. The beta release covers those functionalities defined by the minimal viable product definition of the those components

For this beta release the PIDMR was designed based on the following identified specifications and requirements collected from different stakeholders:

- Implementing a single scalable PID resolving infrastructure for resolving different types of PIDs
- Recognizing and resolving capability for various PIDs including ARK, arXiv, Handle, DOI, SWHID, NBN:URN:DE, BNB:URN:FI and Zenodo PID
- Ability to determine the structure of the information that can be expected from a resolution request
- Supporting the schema and the structure of the information a PID resolution system supports
- Implementing Handle Software and Global Handle Resolving (GHR) Network to develop libraries and tools to integrate PID providers via the GHR

The Test-Bed on which the developing PIDMR is currently running is configured as shown in Code 1.

```
Test bed
OS: Ubuntu 22
4 CPU cores
4GB RAM
40GB Hard disk
```

Code 1

For the PIDMR Backend as given above Handle.Net® Software [3] was used to set up a resolving handle proxy server which is accessible at [4].

For testing the PIDMR a local user interface (UI) was implemented based on the Handle.Net® Software UI on the above

```
<namespace>
<template delimiter="@">
<foreach>
<if value="type" test="equals" expression="URL">
<if value="extension" test="matches" expression="^[a-zA][r|R][X|x][i|i][V|V]:\d{2}(\9|0)[1-9][1[0-2]]\.\d{4,5}(\vd+)?)"?metadata$"
parameter="x">
<value data="http://export.arxiv.org/oai2 verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:arXiv:org:${x[2]}" />
</if>
</if>
</foreach>
</template>
</namespace>
```

Code 2

endpoint at [4].

Based on different types of content which a PID provider may offer three types of resolving modes were defined upon which a resolving matrix was derived and implemented including landing-page-mode, metadata-mode and resource-mode. Table 1 shows the resolving matrix.

	LANDING PAGE	METADATA	RESOURCE
ARK	x	x	-
arXiv	x	x	x
Handle	x	x	-
DOI	x	x	-
SWHID	x	x	x
DE-URN	x	x	x
FI-URN	x	-	-
zenodo PID	x	x	x

Table 1: Resolving Matrix

For resolving different types of PIDs Handle.Net® Software offers a build-in resolving mechanism which make use of a so called Template Handle. An example of the implemented Template Handle for resolving arXiv PID [5] in order to get the metadata of an object is given in Code 2.

For the purpose of resolving mechanism using the Template Handle to work a Handle Prefix has to be registered with the Global Handle Resolving (GHR) Network. In the case of the PIDMR Beta Release 21.T11973/MR was registered as the prefix. To create a resolvable PID the Template Handle uses a delimiter which by default is defined to be @. Using the delimiter an example of a resolvable arXiv PID could look like shown in Code 3.

```
21.T11973/MR@arXiv:2302.00338?metadata
```

Code 3

Here are 21.T11973/MR the Handle Prefix, arXiv:2302.00338 the arXiv PID and metadata the resolving mode. The part arXiv:2302.00338?metadata is called the extension and used in the Template Handle to check against the implemented regex pattern to recognize the PID type. After recognizing, the request is redirected to the given endpoint configured in the data parameter of the value tag given in the Template Handle, thus retrieving the metadata of the object. An API call would then read as given in [6].



# Projektstart zu einheitlichen Filmidentifikatoren in einem Verbundsystem

**Text und Kontakt:**  
Dr. Sven Bingert  
sven.bingert@gwdg.de

Zum 1. November 2023 ist das DFG-geförderte Projekt zur Erstellung eines automatisierten Verbundsystems für audiovisuelle Bestände über einheitliche Filmidentifikatoren (AV-EFI) gestartet. Im Projekt AV-EFI, an dem auch die GWDG beteiligt ist, wird ein institutionsübergreifendes, webbasiertes System entwickelt, bei dem Persistente Identifikatoren (PIDs) eine zentrale Rolle spielen. Dieses System wird dazu dienen, Filme, ihre Fassungen und Datenobjekte aus heterogenen Sammlungen zu identifizieren und zu verknüpfen.

## EINLEITUNG

Im Projekt „Automatisiertes Verbundsystem für Einheitliche Filmidentifikatoren“ (AV-EFI) arbeiten filmhaltende Institutionen mit Infrastrukturanbietern zusammen, um ein zentrales PID-basiertes Verbundsystem zu etablieren. Die Vernetzung von Metadaten aus verschiedenen Quellen ist ein allgemeines Ziel, das nun mit dem Projekt im Bereich der audiovisuellen Bestände erreicht werden soll. Durch den Aufbau eines solchen Verbundsystems werden nicht nur Film- und Medienwissenschaften angesprochen, sondern es wird auch den interdisziplinären Geistes- und Kulturwissenschaften ein neuer Zugang zum kulturellen Erbe unserer Gesellschaft ermöglicht. Die Projektpartner sind das Leibniz-Informationszentrum für Technik und Naturwissenschaften (TIB) als Konsortialleitung, die Stiftung Deutsche Kinemathek – Museum für Film und Fernsehen (SDK), die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) und das Filmmuseum der Landeshauptstadt Düsseldorf (FMDU). Das Projekt wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert und hat eine Laufzeit von November 2023 bis Oktober 2025.

## ZIELSETZUNG

Durch offene Metadaten, die zentral in der PID-Infrastruktur gesichert werden, sollen Filmdaten besser auffindbar und in weitere (Forschungs-)Prozesse integrierbar sein. Das Projekt möchte den Nutzerkreis der (Meta-)Daten erweitern und z. B. durch Workshops filmhaltende Institutionen und Vertreter\*innen verschiedener Fachcommunities schon früh mit einbeziehen. Die dabei erarbeiteten Standards führen zu qualitativen Metadaten, wobei auch Verknüpfungen mit externen Quellen wie z. B. der Normdatenbank der GND (Gemeinsame Normdatei) integriert werden.

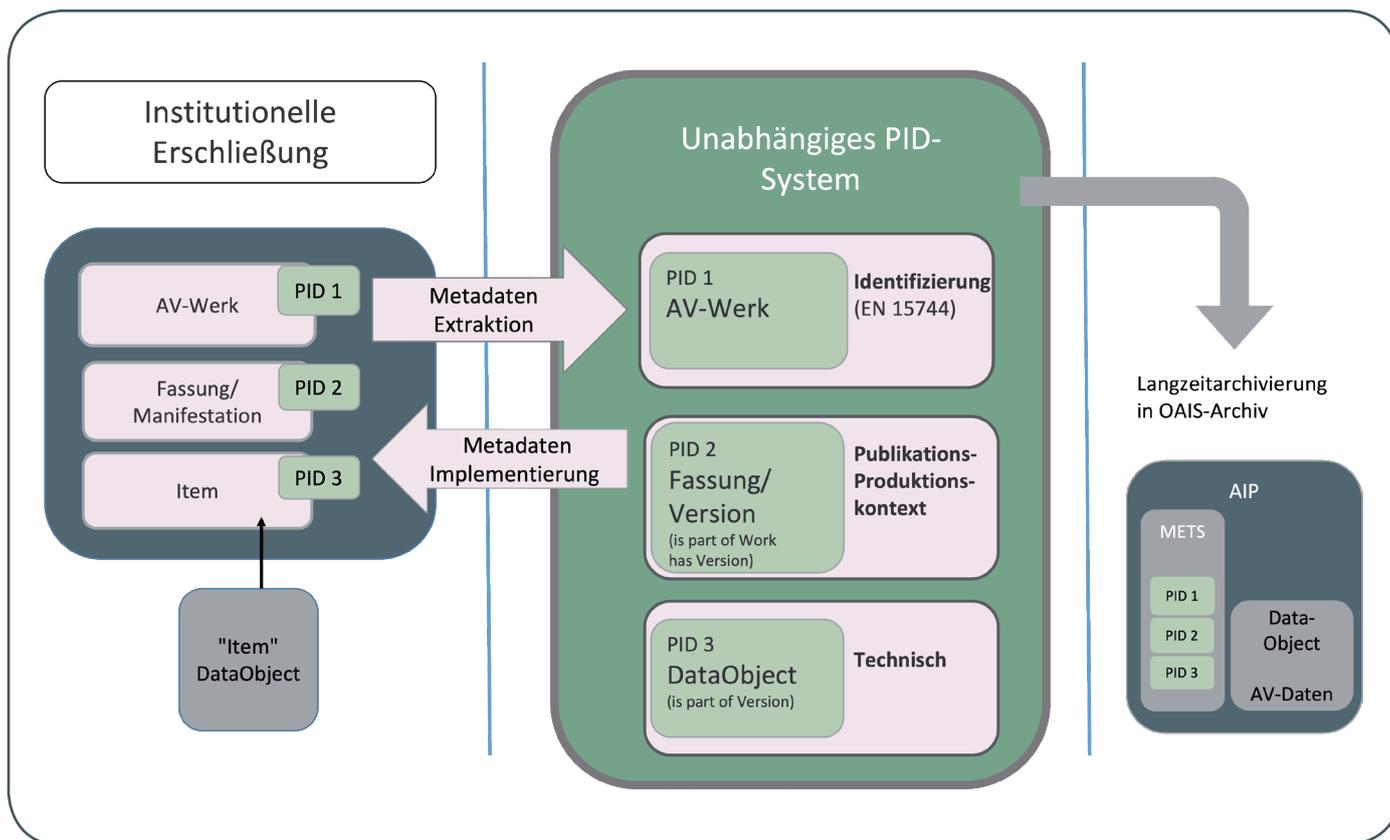
Neben den Standards gilt es dann auch die notwendigen Prozesse zu etablieren. Um die Qualität der in das PID-System eingereichten Metadaten zu gewährleisten, wird eine Validierung vorgeschaltet. Dieser Validierungsschritt verwendet die Type Registry



und Type-API, die aktuell im EU-Projekt FAIRCORE4EOSC [1] als EOSC Core Service implementiert werden. Beide Dienste sind schon seit längerem in Betrieb und werden von der GWDG im Rahmen des Persistent Identifier Consortium for eResearch (ePIC) [2] bereitgestellt. Im nächsten Schritt soll eine automatisierte Prüfung zur Disambiguierung mit modernen NLP-Methoden (NLP = Neuro-Linguistische Programmieren) stattfinden. Die dann final im PID-System registrierten Metadaten sollen dann über einfache Schnittstellen (z. B. REST) für alle Interessierten bereitgestellt werden. Die PIDs können dann verwendet werden, um die Metadaten in institutionellen Datenbanken oder Archivsystemen anzureichern. Abbildung 1 veranschaulicht die zentrale Rolle des PID-Systems für die audiovisuellen Materialien.

## Start of the AV-EFI Project

The DFG-funded project to create an automated network system for audiovisual collections using uniform film identifiers (AV-EFI) was launched on November 1, 2023. In the AV-EFI project, in which the GWDG is also participating, a cross-institutional, web-based system is being developed in which Persistent Identifiers (PIDs) play a central role. This system will be used to identify and link films, their versions and data objects from heterogeneous collections.



1\_Austausch von Metadaten und Verwendung von PIDs im Verbundsystem

## ARBEITSPAKETE

Die Arbeiten im Projekt AV-EFI sind in vier Arbeitspakete (AP) untergliedert. Das Arbeitspaket 1 kümmert sich um Konzeption, Anforderungsanalyse und Koordination. Dabei steht der Austausch mit den relevanten Stakeholdern im Vordergrund und wird mit verschiedenen Workshops vertieft. Im AP1 werden die Verbundpartner (Antragsteller), Praxispartner und Fachcommunities berücksichtigt. Praxispartner sind filmhaltende Institutionen, die mit AV-Materialien arbeiten und Metadaten als Nachweis zur Verfügung stellen, wobei Fachcommunities die Forschungsperspektive abdecken. Für die Integration der Fachcommunities wurde schon zum Projektantrag mit zehn Institutionen bzw. Wissenschaftler\*innen in Deutschland eine konkrete Vereinbarung zur Zusammenarbeit (Letter of Intent) getroffen.

Als Teil der Konzeption werden verschiedenen Richtlinien bezüglich der Zusammenarbeit erarbeitet. Ziel ist, eine effektive und zielführende Umgebung zu schaffen, in der Unklarheiten und offene Punkte kollaborativ geklärt werden können. Ein weiterer Aspekt ist der Betrieb der zu entwickelnden AV-EFI-Plattform über die Projektlaufzeit hinaus. Wie kann neben der Persistenz der Identifikatoren auch die weitere Nutzung durch die unterschiedlichen Interessengruppen gewährleistet werden.

Das Arbeitspaket 2 fokussiert sich auf die Bestände und die Metadaten. Neben der Auswahl geeigneter Bestände werden die Konzeption zu Fassung oder Gattung definiert und eine Schemaerweiterung diskutiert. Ein initiales Schema V1.0 wurde in einem Vorprojekt erarbeitet und auf GitHub veröffentlicht [3]. Eine Anpassung des Schemas erfolgt dann nach den im AP1 festgelegten Richtlinien. Ein weiterer Aspekt des AP2 ist die Erstellung der Konzeption eines Abgleichs und der Definition von Ähnlichkeitsparametern

und Schwellenwerten, um eine Disambiguierung und Deduplizierung im PID-System zu ermöglichen.

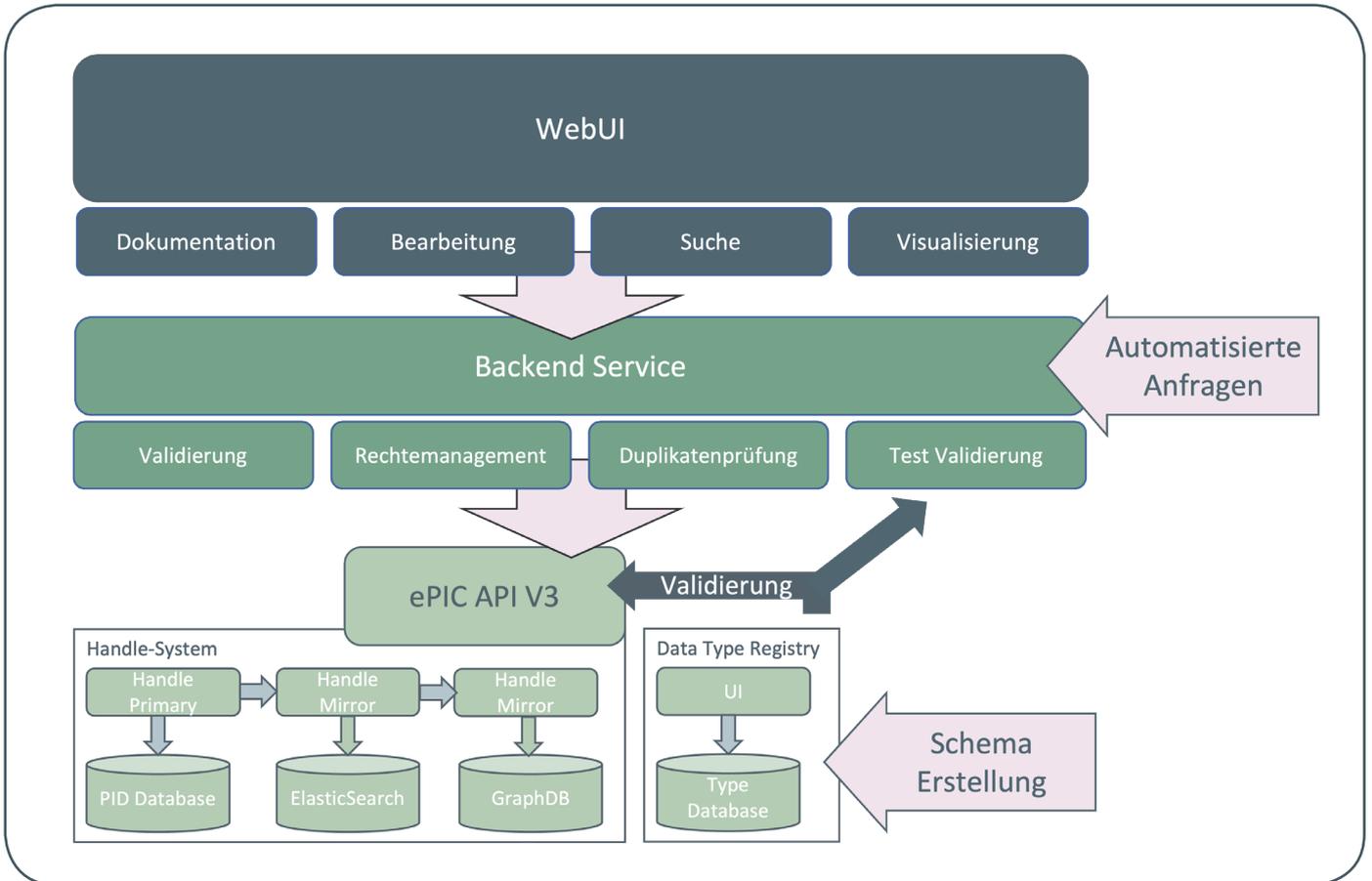
Im dritten Arbeitspaket (AP3) werden die technischen Grundlagen entwickelt und bereitgestellt. Dazu zählt das PID-System für die Erstellung und Verwaltung von PIDs in einem festen Namensraum. Ergänzend dazu werden verschiedene erweiterte Dienste entwickelt, welche die Anforderungen aus AP1 und AP2 umsetzen (siehe Abbildung 2).

Diese Dienste werden nach modernen Standards der Softwareentwicklung erstellt und nach den Vorgaben der GWDG als Infrastrukturbetreiber in ihr Dienstportfolio integriert. Das bedeutet auch, dass Vorgaben zum Datenschutz und zur IT-Sicherheit Teil des AP3 sind.

Im Arbeitspaket 4 wird das Web-User-Interface entwickelt, das neben der automatisierten Schnittstelle zum Backend eine weitere Möglichkeit zur Interaktion mit dem PID-System bietet. Dieses GUI stellt die Möglichkeit der redaktionellen Bearbeitung von inhaltlichen Konflikten zur Verfügung. Autorisierte Nutzer\*innen sollen damit, nach den in AP1 festgelegten Richtlinien, Entscheidungen z. B. zur Deduplizierung treffen, wenn diese vom System nicht automatisch gelöst werden können.

## AUSBLICK

Das zu entwickelnde Verbundsystem wird in seiner Gesamtheit technisch wie organisatorisch erst konzipiert und initial aufgebaut. Zum Laufzeitende soll es anhand bereits beschriebener und weiterer Anwendungsfälle, Anforderungsanalyse und Aktivitäten aller Partner soweit entwickelt sein, dass es bereits sinnvoll funktioniert und es eine verbindliche Perspektive für Weiterentwicklung und nachhaltigen Betrieb geben wird. Das technische Herzstück



2\_Darstellung der Diensteebenen und deren Aufgaben. Das Web-User-Interface, entwickelt in AP4, kommuniziert mit dem Backendsystem, das wiederum die Schnittstelle zum PID-System darstellt.

des Verbundsystems, das PID-System mit den Basisdiensten, wird dann bereits im dauerhaft verfügbaren Produktivbetrieb sein. Das langfristige Ziel der Projektpartner ist, neben der Integration einer großen Anzahl an filmhaltenden Institutionen in das Verbundsystem, die Etablierung des einheitlichen Filmidentifikators (EFI) auch im internationalen Kontext.

#### LINKS

- [1] <https://faircore4eosc.eu>
- [2] <https://www.pidconsortium.net>
- [3] <https://github.com/AV-EFI/av-efi-schema>
- [4] Projektleitung bzw. Webseite des Projekts: <https://www.tib.eu/de/forschung-entwicklung/projektuebersicht/projektsteckbrief/av-efi>





# MS SharePoint

## KOLLABORATION LEICHT GEMACHT!

### Ihre Anforderung

Sie möchten eine kooperative Kommunikations- und Informationsplattform für Mitarbeiter\*innen einrichten, die ständig und von überall verfügbar ist. Sie benötigen ein integriertes Dokumentenmanagementsystem und möchten gemeinsame Besprechungen und Termine planen und verwalten.

### Unser Angebot

Wir bieten Ihnen SharePoint als Kollaborationsplattform. Wir können z. B. eine SharePoint Site Collection als gemeinsames Portal für Ihre Arbeitsgruppe oder Ihr Projektteam einrichten. Eine solche Site Collection kann sowohl in Englisch als auch in Deutsch präsentiert werden. Mit einer umfangreichen Auswahl an Schablonen, Apps und Layout-Vorlagen können Sie das Design Ihrer Site Collection anpassen. Der Zugriff erfolgt über GWWDG-Benutzerkonten. Weitere Authentifizierungsverfahren sind möglich.

### Ihre Vorteile

- > Einheitliches Dokumenten-Managementsystem
- > Umfangreiche Listen und Bibliotheksfunktionen für Dokumente, Bilder oder Dateien

- > Steigern der Produktivität der Mitarbeiter\*innen durch vereinfachte tägliche Geschäftsaktivitäten.
- > Einfaches Planen und Protokollieren von Besprechungen
- > Führen nicht öffentlicher Diskussionsrunden
- > Wissensmanagement: Aufbau eines Wikis für Ihre Mitarbeiter\*innen
- > Bereitstellung von Informationen und Fachwissen für Mitarbeiter\*innen
- > Geringer Entwicklungs- und Pflegeaufwand der SharePoint-Plattform für Benutzer\*innen
- > Individuell anpassbares Layout und Design
- > Optimale MS Office-Anbindung
- > Einfache Benutzer- und Gruppenverwaltung

### Interessiert?

Der Dienst steht allen Mitgliedern der Max-Planck-Gesellschaft und der Universität Göttingen zur Verfügung. Voraussetzung für die Nutzung ist die Benennung eines Ansprechpartners, der die Administration Ihrer Site Collection übernehmen soll. Wenn Sie SharePoint nutzen möchten, senden Sie bitte eine entsprechende E-Mail an [support@gwdg.de](mailto:support@gwdg.de). Nähere Informationen zu SharePoint sind auf der u. g. Webseite zu finden.

# Start der „Landesinitiative Forschungsdatenmanagement Niedersachsen“

**Text:**  
Pressemitteilung von  
Hochschule.digital Niedersachsen  
vom 27.11.2023 mit Ergänzungen

Mit 15,7 Millionen Euro fördert das Niedersächsische Ministerium für Wissenschaft und Kultur (MWK) den Aufbau der „Landesinitiative Forschungsdatenmanagement Niedersachsen“. Damit werden Empfehlungen der Wissenschaftlichen Kommission Niedersachsen umgesetzt. Mit einem „Kick-off“ am 24. November 2023 ist das Verbundprojekt, an dem alle niedersächsischen Hochschulen mitwirken, gestartet. Die neue Landesinitiative ist Teil von *zukunft.niedersachsen*, dem gemeinsamen Programm von MWK und VolkswagenStiftung zur Stärkung von Wissenschaft und Technik in Forschung und Lehre. Den organisatorischen Rahmen des Projekts bildet der landesweite Digitalisierungsverbund *Hochschule.digital Niedersachsen*.

Die neue Landesinitiative hat zum Ziel, das Forschungsdatenmanagement in Niedersachsen in die Breite der Wissenschaft zu tragen und ein flächendeckendes Angebot zur Unterstützung der Forschenden zu schaffen. Auf diese Weise kann die Qualität im Forschungsprozess durch bewussten und kompetenten Umgang mit Forschungsdaten stetig erhöht werden.

Ein dreiteiliges Förderprogramm dient dazu, eine übergreifende Struktur und Strategie des Forschungsdatenmanagements in Niedersachsen zu entwickeln und trägt zugleich den unterschiedlichen Rahmenbedingungen und Ausprägungen an den Hochschulen Rechnung. Neben der Etablierung einer zentralen Service- und Beratungsstelle, die in Form eines verteilten Netzwerkes Unterstützungsangebote weiterentwickelt und allen Hochschulen zur Verfügung stellt (Säule 1), werden Ressourcen zum Aus- und Aufbau von Basisfähigkeiten im Forschungsdatenmanagement für jene Hochschulen bereitgestellt, die bislang noch kein grundlegendes Service-Angebot an Forschungsdatenmanagement eingerichtet haben (Säule 2). Über einen Projektfonds können darüber hinaus besonders innovative Projekte gefördert werden, die ihre Wirkung über die eigene Einrichtung hinaus entfalten und damit der Verbesserung des Forschungsdatenmanagements im Land dienen (Säule 3).

Der Ausbau einer leistungsfähigen, effizienten und innovativen Forschungsdateninfrastruktur ist eine entscheidende Voraussetzung für den Erfolg der niedersächsischen Hochschulen im nationalen und internationalen Wettbewerb, für die Forschung und insbesondere für die Spitzenforschung und Exzellenz.

„Die Währung der Wissenschaft sind wissenschaftliche Daten. Aus den Daten kann neues Wissen entstehen, welches auch über die Fächer hinweg nützlich ist, um Lösungen für die großen gesellschaftlichen Fragen zu finden. Die Landesinitiative Forschungsdatenmanagement unterstützt die Hochschulen, Daten

auffindbar, bedarfsgerecht und qualitätsgesichert allen Forschenden bereitzustellen“, so Prof. Dr. Joachim Schachtner, Staatssekretär für Wissenschaft und Kultur.

Prof. Dr. Norbert Lossau, seit 01.07.2023 Gründungsdirektor der *Hochschule.digital Niedersachsen*, unterstreicht den kooperativen Ansatz des Projektes: „Die Landesinitiative fördert die standortübergreifende Zusammenarbeit der Hochschulen. Zugleich zielen die geplanten Maßnahmen auf die Anbindung an die Nationale Forschungsdateninfrastruktur (NFDI).“

Die Initiative legt den Grundstein für den Aufbau einer gemeinschaftlichen, niedersachsenweiten Struktur und Strategie des Forschungsdatenmanagements.

„Nach mehrjährigen Vorarbeiten ist es den niedersächsischen

## Launch of the “Lower Saxony Research Data Management Initiative”

The Lower Saxony Ministry of Science and Culture (MWK) is providing 15.7 million euros in funding to set up the “Lower Saxony Research Data Management Initiative”. This will implement the recommendations of the Lower Saxony Scientific Commission. The joint project, in which all universities in Lower Saxony are involved, was launched with a “kick-off” on November 24, 2023. The new state initiative is part of *zukunft.niedersachsen*, the joint program of the MWK and the Volkswagen Foundation to strengthen science and technology in research and teaching. The state-wide digitalization network *Hochschule.digital Niedersachsen* forms the organizational framework of the project.



v. l. n. r.: Prof. Dr. Josef von Helden (Hochschule Hannover, Sprecher der Ständigen LHK-Kommission für Digitalisierung), Prof. Dr. Wolfgang Viöl (Hochschule für angewandte Wissenschaft und Kunst Hildesheim/Holzmin-den/Göttingen), Prof. Dr. Joachim Schachtner (Staatssekretär für Wissen-schaft und Kultur), Robert Strötgen (Technische Universität Braunschweig, Sprecher der LHK-AG Forschungsdatenmanagement), Prof. Dr. Norbert Lossau (Gründungsdirektor der Hochschule.digital Niedersachsen)

Hochschulen gelungen, in einem gemeinsamen Konzept auf die heterogene Lage beim Forschungsdatenmanagement einzugehen und dabei die Empfehlungen der Wissenschaftlichen Kommission Niedersachsen konstruktiv aufzunehmen. Wir freuen uns sehr auf den Start der Landesinitiative!“ sagt Robert Strötgen, Direktor der Universitätsbibliothek der TU Braunschweig und gemeinsam mit Dr. Janna Neumann (Technische Informationsbibliothek) Sprecher der niedersächsischen AG Forschungsdatenmanagement, an der alle Hochschulen beteiligt sind.

## Hintergrund

### Forschungsdaten und Forschungsdatenmanagement in Niedersachsen

Forschungsdaten bilden sowohl die Grundlage als auch das Ergebnis wissenschaftlichen Arbeitens. Ihre langfristige Sicherung und Bereitstellung leistet einen Beitrag zur Nachvollziehbarkeit und Qualität wissenschaftlicher Arbeit und eröffnet wichtige Anschlussmöglichkeiten für die weitere Forschung. Die mit der Etablierung von Digitalisierungsprozessen einhergehende Zunahme der Datenintensität in vielen wissenschaftlichen Disziplinen stellt dabei neue Anforderungen an die Hochschulen.

Im Oktober 2021 hatte eine von der Wissenschaftlichen Kommission Niedersachsen eingesetzte Arbeitsgruppe eine Analyse zum Umsetzungsstand des Forschungsdatenmanagements in Niedersachsen vorgelegt und mit Empfehlungen an die Hochschulen und das Land verbunden. In der Folge setzte die Ständige Kommission für Digitalisierung der Landeshochschulkonferenz Niedersachsen eine Arbeitsgruppe zum Forschungsdatenmanagement ein, um einen Plan zur operativen Umsetzung der Empfehlungen zu erarbeiten. An diesem Prozess waren Vertreterinnen und Vertreter aller 20 niedersächsischen Hochschulen in staatlicher Trägerschaft beteiligt.

### Hochschule.digital Niedersachsen

Die Dachinitiative *Hochschule.digital Niedersachsen* ist eine

gemeinsame Gründung der Landeshochschulkonferenz Niedersachsen, des Niedersächsischen Ministeriums für Wissenschaft und Kultur sowie der VolkswagenStiftung, der alle niedersächsischen Hochschulen in staatlicher Verantwortung angehören. Mit dem landesweiten Digitalisierungsverbund wird eine tragfähige gemeinsame und nachhaltige Digitalisierungsstrategie und -struktur für die niedersächsischen Hochschulen entwickelt, die dazu dient, digitale Technologien noch breiter und professioneller in Studium und Lehre, Forschung sowie Verwaltung einzusetzen.

### *zukunft.niedersachsen*

Im Februar 2023 wurde das Förderprogramm *zukunft.niedersachsen* durch Wissenschaftsminister Falko Mohrs, Dr. Georg Schütte, Generalsekretär der VolkswagenStiftung, und Prof. Dr. Susanne Menzel-Riedl, Vorsitzende der Landeshochschulkonferenz, vorgestellt. Durch die Sonderdividende aus dem Börsengang der Porsche AG stehen in den kommenden Jahren rund 576,3 Millionen Euro zusätzlich für die Förderung von Wissenschaft und Technik in Forschung und Lehre zur Verfügung. Im Mittelpunkt des Programms stehen die drei Zukunftsfelder Transformation, Digitalität und Spitzenforschung.

## Kontakt

### „Landesinitiative Forschungsdatenmanagement Niedersachsen“

- Robert Strötgen, Technische Universität Braunschweig: [r.stroetgen@tu-braunschweig.de](mailto:r.stroetgen@tu-braunschweig.de)
- Dr. Janna Neumann, TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften und Universitätsbibliothek: [Janna.Neumann@tib.eu](mailto:Janna.Neumann@tib.eu)

### Hochschule.digital Niedersachsen

- Krystyna Redeker-Weißer, Koordinierungsstelle: [krystyna.redeker-weisser@hd-nds.de](mailto:krystyna.redeker-weisser@hd-nds.de)
- Informationen zur Dachinitiative *Hochschule.digital Niedersachsen*: <https://www.hochschuledigital-niedersachsen.de>

## Göttinger Beteiligung an der Landesinitiative

Die GWDG war während der Planungs- und Antragsphase Mitglied einer Kerngruppe, die den Antrag federführend begleitet und auf den Weg gebracht hat. Das nun erfolgte „Kick-Off“ ist zum einen der erfolgreiche Abschluss dieser Vorbereitungsphase und zum anderen der Beginn der nicht minder herausfordernden Umsetzungsphase der ambitionierten Pläne der Landesinitiative. In enger Zusammenarbeit mit den Partnern in Niedersachsen, aber auch in Kollaboration mit der deutschen Nationalen Forschungsdateninfrastruktur (NFDI) sowie europäischen Partnern, wird die GWDG als Partner der Göttingen eResearch Alliance die Landesinitiative voranbringen. Gemeinsam mit der Niedersächsischen Staats- und Universitätsbibliothek Göttingen (SUB) wird die GWDG ihre ganze Erfahrung von mehr als zehn Jahren Forschungsdatenmanagement in den Aufbau von Beratungs- und Trainingsangeboten, den Ausbau des Service-Portfolios sowie die Etablierung einer dauerhaften Governance-Struktur einbringen. ●



# Data Science Summer School 2023

## Text and Contact:

Dr. Sven Bingert  
sven.bingert@gwdg.de

Lena Steilen  
lena.steilen@gwdg.de

This year, after a long break due to the coronavirus pandemic, the Data Science Summer School finally took place again in Göttingen. The 30 participants from twelve different countries experienced a varied and challenging program on various data science topics. Over twelve days, they learnt about deep learning, but also took a critical look at the topics of security, AI safety and reproducible research. The many positive responses show the success of the Summer School series and motivate us to continue the concept in the coming years.

## INTRODUCTION

This year's Data Science Summer School (DS3), September 11th – September 22nd, was something special in many respects. Not only was it the first Summer School after the pandemic-related break, but the event was also closely linked to two other events within the HeKKSaGOn network (<https://www.hekksagon.net>).

Göttingen was selected to host a HeKKSaGOn Symposium on AI and the official HeKKSaGOn President's meeting 2023. Both events took place in parallel to the Summer School program. The Summer School benefited from the fact that experts on AI could be motivated to come to Göttingen and speak at the symposium or teach at the Summer School. At the end of the program, the participants of the Summer School had to present their results and their learnings at the President's Meeting. This was particularly important as the promotion of young scientists is a high priority in the objectives of the HeKKSaGOn network. At the Conference Dinner of the Presidents Meeting the expert delegations and the participants met jointly at the new "Göttinger Rechenzentrum" to get in touch.

The organisation was particularly challenging this year, not only because of the parallel events that took place in a similar context. In general, it was clear early on that a large number of events

would be taking place in Göttingen and the surrounding area during this period as many fairs, conferences and meetings should take place face to face finally again. This made it difficult to find suitable accommodation for the participants, which in the end led us to the decision to officially start the Summer School one day later.

As September approached, the necessary and sensible bomb disposal at the "Schützenplatz" was announced. This event then had a direct impact on hotel bookings and train connections, which

## Data Science Summer School 2023

Nach einer langen Pause aufgrund der Coronavirus-Pandemie fand die Data Science Summer School in diesem Jahr endlich wieder in Göttingen statt. Die 30 Teilnehmer\*innen aus zwölf verschiedenen Ländern erlebten ein abwechslungsreiches und anspruchsvolles Programm zu verschiedenen Data Science-Themen. In zwölf Tagen lernten sie über Deep Learning, setzten sich aber auch kritisch mit den Themen Sicherheit, KI-Sicherheit und reproduzierbare Forschung auseinander. Die vielen positiven Rückmeldungen zeigen den Erfolg der Sommerschulreihe und motivieren uns, das Konzept auch in den kommenden Jahren fortzusetzen.



ultimately led to the Summer School being shortened by half a day at the end in order to allow the participants a practicable return journey. Finally, of course, it should be mentioned that the evacuation was successful and all participants were able to return home safely.

## SCIENTIFIC PROGRAM

The program aimed to give the participants insights in the multi-layered topics and aspects of Data Science. Knowledge and skills in Data Science are indispensable for researchers of almost every discipline nowadays. Researchers profit from knowing about the scope and application areas of Data Science in daily scientific work and benefit from understanding the belonging implications for scientific inquiries, and the skills required by a data scientist to be productive in a world deluged by data.

The various lectures gave insights on the multi-layered topics and methods belonging to Data Science. Such as Data Quality, Exploratory and Quantitative Data Analysis, Deep Learning, Application examples, Quantum Machine Learning AI Safety and Security and Scalable AI.

Due to the stroke of luck that the two other HeKKSaGOn events took place in Göttingen as well, it was possible to involve lecturers from the Network (see program).

The program began with some warm words of welcome from Prof. Dr. Hiltraud Casper-Hehne, Central Internationalization Officer of the University of Göttingen, followed by the introduction to Data Science and the first session of the Deep Learning course. Deep Learning was the focus of this year's school. The participants then introduced themselves in a number of flash talks or short presentations. These presentations are the best icebreakers and show the potential, skills and enthusiasm of the speakers.

The following days began with another session on Deep Learning and then covered various relevant areas, such as statistics or metadata. In addition to the technical topics, the program also addressed social and ethical issues. These include reproducible science as well as the current dynamics of the development of the capabilities of AI models.

We were also pleased to be able to present another important aspect of the rapid technological development with a course on Quantum Machine Learning. The lecturers of the Deep Learning course set up a challenge over a few days and the participants had fun in solving the task(s). The winners got a University of Göttingen T-Shirt and were celebrated in the group.



## SOCIAL PROGRAM

One aspect of the Summer School is the international networking of young scientists from various disciplines. For this reason, various social activities were offered, which were gratefully accepted. During communal barbecues and hikes, discussions centred on both specialist and intercultural topics. The city of Göttingen and its long tradition in science could also be admired during the city tours.

## CONCLUSION

Both the participants and the organisers were completely satisfied with the Summer School. The GWDG, or rather the new data centre, was a good choice as an event location. The participants were able to actively socialise with each other and with the speakers during the breaks in the attractive foyer. Lunch was taken in the lunchbox of the "Studentenwerk Göttingen" just in five minutes walking distance. And for the daily journey between the city centre and the new data centre, participants were provided with bus tickets from GoeVB. Once again, we would like to thank the "Studentenwerk Göttingen" and the "GoeVB" for their excellent cooperation and support.

The participants showed a high level of social interaction during group work as well as during breaks and social events. The participants made a short video at the end of the Summer School that captures the atmosphere very well ([https://www.youtube.com/watch?v=\\_6nC96zkl-U&t=57s](https://www.youtube.com/watch?v=_6nC96zkl-U&t=57s)). Or to cite one of the participants Agresa Qosja: "Our cohort represented a global mosaic, with students hailing from Albania, Germany, Austria, Egypt, Morocco, Japan, China, Indonesia, Pakistan, Bangladesh, India, Ghana, and more. The cultural richness added a unique dimension to our learning." ■



# GWDG Pad

## KOLLABORATION LEICHT GEMACHT!

### Ihre Anforderung

Sie möchten allein oder gemeinsam mit Ihrem Team unkompliziert an Textdokumenten arbeiten oder Präsentationen erstellen und dabei auf eine Vielzahl nützlicher Funktionen zurückgreifen. Ihre Änderungen sollen sowohl für Sie als auch Ihre Teamkolleg\*innen direkt und in Echtzeit einsehbar sein. Sie möchten die Lese- oder Schreiberechtigung für Ihre Dokumente einschränken können, sodass Sie Ihre Daten vor unbefugtem Zugriff schützen können. Zudem wollen Sie in der Auswahl Ihrer Endgeräte flexibel sein, sowohl mobiler Zugriff als auch Desktop-Varianten sollen unterstützt werden.

### Unser Angebot

Auf Basis der freien Software „HedgeDoc“ bieten wir Ihnen einen Dienst, mit dem Sie schnell und unkompliziert Dokumente erstellen, mit anderen Personen teilen und gemeinsam bearbeiten können.

### Ihre Vorteile

- > Kollaborativer Echtzeit-Editor
- > Übersicht über alle Ihre Dokumente nach Login

- > Unterstützung von UML-Diagrammen, mathematischen Formeln, Syntax-Highlighting, Musiknoten und vielem mehr
- > Modi zum Erstellen und Vorführen von Präsentationen
- > Einbinden externer Ressourcen wie Videos, PDF-Dateien oder SlideShare
- > Autovervollständigungs-Funktion für Markdown-Ausdrücke
- > Zugriffsbeschränkungen für jedes Dokument einstellbar
- > Veröffentlichung von Dokumenten möglich
- > Webbrowser ausreichend zur Benutzung, keine weitere Installation von Software nötig

### Interessiert?

Wenn Sie unseren Dienst „GWDG Pad“ unter <https://pad.gwdg.de> nutzen möchten, benötigen Sie lediglich einen aktuellen Webbrowser. Um eigene Dokumente erstellen zu können, ist zusätzlich die Verwendung eines gültigen GWDG-Accounts oder die einmalige Registrierung unter <https://www.gwdg.de/registration> erforderlich.

# Nationales Hochleistungsrechnen – Vorhang auf zur 1. NHR-Konferenz 2023

## Text und Kontakt:

Alexander Goldmann  
alexander.goldmann@gwdg.de

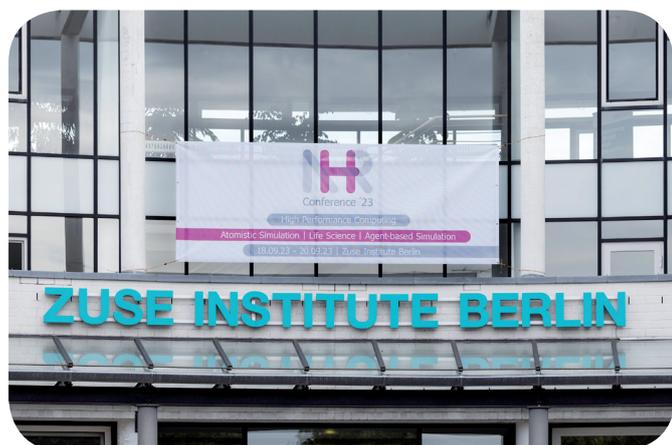
Gelungene Premiere: Vom 18. bis 20. September 2023 fand am Zuse-Institut Berlin die 1. NHR-Konferenz statt. Sie richtete sich an den ersten beiden Tagen im wissenschaftlichen Teil besonders an die externe HPC-Anwendergemeinschaft und wurde vom Zielpublikum gut angenommen. Die diesjährigen Themenschwerpunkte lagen in den Bereichen Atomistische Simulation, Lebenswissenschaften und Agentenbasierte Simulation. Auch die GWDG war mit mehreren Beiträgen am abwechslungsreichen Programm beteiligt. Die Konferenzschwerpunkte werden jährlich wechseln und sich auf verschiedene wissenschaftliche Themen konzentrieren.

## VIELE BEITRÄGE UND EIN INTENSIVER AUSTAUSCH

Mit hochkarätigen Keynote-Speakern und weiteren Vorträgen zu den diesjährigen wissenschaftlichen Themen bot die Konferenz die passende Grundlage für einen intensiven wissenschaftlichen Austausch über Fachgrenzen hinweg. Durch verschiedene Vorträge und Präsentationen von Postern hatten Nutzer\*innen und auch Beratungs- und Betriebsteams der neun NHR-Zentren (NHR = Nationales Hochleistungsrechnen) die Gelegenheit, ihre Forschung mit bekannten HPC-Expert\*innen zu diskutieren.

Seitens der GWDG wurden unter anderem Themen wie die „Verwendung von SecureHPC zur Verarbeitung sensibler MRI-Gesundheitsdaten“ oder „Groß angelegte synthetische Waldpunktwolkenerzeugung auf HPC-Systemen“ vorgestellt und darüber ausgiebig ausgetauscht. Es gab viele Diskussionen zwischen

den Nutzer\*innen und den Mitarbeiter\*innen der NHR-Zentren. In den verschiedenen Breakout-Sessions zu den Themengebieten „Numerische Strömungsmechanik“, „Digitale Geisteswissenschaften“, „Lebenswissenschaften“ und „Physik/Chemie“ wurde gut erkennbar, dass die Teilnehmer\*innen sehr unterschiedliche Erfahrungen aus der Sicht von Administrator\*innen, Entwickler\*innen, Doktorand\*innen und Professor\*innen einbringen konnten. So kristallisierte sich beispielsweise in der von Patrick Höhn (GWDG) geleiteten Session zur numerischen Strömungsmechanik heraus, dass bislang kein wirkliches Forum der CFD-Fachberater:innen (CFD = Computational Fluid Dynamics) im NHR vorhanden ist. Dieses soll nun im Laufe des kommenden Jahres aufgebaut werden. Die Ergebnisse bei der Vertiefung der Zusammenarbeit sollen dann bei der nächsten 2. NHR-Konferenz vorgestellt werden.



1\_Der Veranstaltungsort der 1. NHR-Konferenz

## National High Performance Computing (NHR) – Curtain up for the 1st NHR Conference 2023

Successful premiere: The 1st NHR Conference took place at the Zuse Institute Berlin from September 18 to 20, 2023. The scientific part of the conference, which was aimed particularly at the external HPC user community on the first two days, was well received by the target audience. This year's main topics were atomistic simulation, life sciences and agent-based simulation. The GWDG was also involved in the diverse program with several contributions. The focus of the conference will change annually and concentrate on different scientific topics.



2\_Teilnehmer\*innen der 1. NHR-Konferenz



3\_Die Vorträge wurden gut besucht

Die von Christoph Hottenroth (GWDG) und Tobias Watermann (ZIB) geführte Admin-Session wurde ebenfalls sehr gut besucht und bot viel Platz für konstruktive Diskussionen. Hier präsentierte Freja Nordsiek (GWDG) interessante Informationen zum Thema „Container“ im HPC-Bereich sowie „Abfangen von rsync/sftp auf Übertragungsknoten aus Sicherheitsgründen“ und Hendrik Nolte (GWDG) referierte über „Warewulf“, ein Computer-Cluster-Implementierungs-Toolkit, das die Installation eines Clusters und die

langfristige Verwaltung erleichtern soll. Die Admin-Session wurde durch Beiträge von Tobias Poschwatta (ZIB) mit dem Thema „Imagebau“ und Nils Winnwa (PC<sup>2</sup>) „Kompakte Präsentation über (Ersatz-)Teile im Bestandsmanagement“ abgerundet.

## AUSBLICK

Zusammenfassend war die 1. NHR-Konferenz ein Erfolg sowohl für die Ausrichtenden als auch die Teilnehmer\*innen. Durch den persönlichen Austausch in einer professionellen Atmosphäre und im Rahmen einer gut organisierten Veranstaltung konnten viele konstruktive Gespräche geführt und ein umfangreiches Networking betrieben werden. Inhaltlich hervorragend konzipierte Vorträge rundeten die zwei Tage ab und unterstützen dabei, NHR-Projekte auch in Zukunft weiter voranzutreiben.

Die 2. NHR-Konferenz wird am 9. und 10. September 2024 in Darmstadt stattfinden und sich auf den inhaltlichen Themenschwerpunkt „Computergestützte Ingenieurwissenschaften“ konzentrieren.

Weitere Informationen zur 1. NHR-Konferenz sind unter dem URL [https://events.uni-paderborn.de/event/487/attachments/99/234/Book\\_of\\_Abstracts\\_NHR\\_Conference\\_2023.pdf](https://events.uni-paderborn.de/event/487/attachments/99/234/Book_of_Abstracts_NHR_Conference_2023.pdf) zu finden. ■

# „Bei Euch wird es auch immer schwieriger?“ – Die 2. IT4Science-Days der Helmholtz-IT und der Max-Planck-IT

**Text und Kontakt:**  
Organisationskomitee  
der IT4Science-Days  
dv-treffen@mpg

Cyberfälle, Mangel an Ressourcen – vor allem Personal – wenig Licht am Horizont; das verbindet die ITler\*innen in den Wissenschaften. Noch mehr vielleicht die unverwüsthliche Liebe zur beharrlichen Suche nach Lösungen. Im Feld der KI zum Beispiel gibt es mittlerweile freie Alternativen. Es wurde viel und lange diskutiert beim 40. DV-Treffen der Max-Planck-Institute im Rahmen der 2. IT4Science-Days vom 26. bis 28. September 2023 gemeinsam mit den Kolleg\*innen der Helmholtz-Gemeinschaft (HGF) am Max-Delbrück-Centrum für Molekulare Medizin in Berlin-Buch. Jetzt müssen noch Management und Wissenschaftler\*innen beitragen und tun, was getan werden muss. Von mancher Zeit merkt man erst im Rückblick, wie wichtig sie war. So eine Zeit ist jetzt.

Mit einem wohlwollenden Grußwort eröffnete die Generalsekretärin der MPG Simone Schwanitz den diesjährigen großen „Opernball“ der IT. Sie unterstrich die Bedeutung des Austausches und der Kooperation zwischen der MPG und dem Gastgeber HGF als außeruniversitäre Forschungseinrichtungen. Der mittlerweile berentete aber immer noch aktive Peter Wittenburg gab im Anschluss einen Rückblick auf die 40-jährige Geschichte des DV-Treffens. Nicht nur sagt man statt DV heute IT – auch vieles andere hat sich sehr verändert, professionalisiert. IT zu kaufen ist nicht schwer, sie zu betreiben aber sehr.

## KEYNOTES

Prof. Dr. Jochem Marotzke, Direktor am MPI für Meteorologie in Hamburg und derzeit BAR-Vorsitzender (BAR = Beratender Ausschuss für Rechenanlagen), hielt am frühen Abend des ersten Tages seine vielbeachtete und richtungweisende Keynote *Gedanken eines Wissenschaftlers zur IT*. Das Fachpublikum war höchst angetan. Er diskutierte die Rolle der IT und des BAR für die Wissenschaft und leitete strategisch wichtige Handlungsbedarfe ab: Mehr Kommunikation und Koordination für die IT – und bessere Bezahlungsmöglichkeiten. Die sind derzeit auf TVöD-Bund E 13 und Gewinnungszulage beschränkt und liegen in der Praxis deutlich darunter.

Chat-KI geht auch ohne „böse“: Die zweite Keynote am folgenden Tag wurde von Frank Karlitschek, dem CEO der Nextcloud GmbH und bekannten Entwickler von Sync-and-Share-Lösungen, gehalten. Es ging um *Digitale Souveränität* und die Möglichkeit,

als Institut oder innerhalb der MPG zeitgemäße und zukunftsweisende Dienste selbst anzubieten, unabhängig(er) von den großen Cloud-Anbietern. Karlitscheks erfreuliche Erkenntnis war, dass dies auch für KI-basierte Funktionalität mit dem *Nextcloud Assistant* seit kurzem möglich sei. Bisher bietet das leider noch keine Einrichtung an.

“It’s getting more and more difficult for you too?” – The 2nd IT4Science Days of the Helmholtz IT and the Max Planck IT

Cyber incidents, lack of resources – especially personnel – little light on the horizon, that’s what unites IT specialists in the sciences. Perhaps even more so is the indestructible love for the persistent search for solutions. In the field of AI, for example, meanwhile there exist alternatives with a free license. There was much and lengthy discussion at the 40th IT meeting of the Max Planck Institutes as part of the 2nd IT4Science Days from September 26 to 28, 2023, together with colleagues from the Helmholtz Association (HGF) at the Max Delbrück Center for Molecular Medicine in Berlin-Buch. Now management and scientists still have to contribute and do what needs to be done. Sometimes you only realize how important a time it was when you look back. Now is such a time.



## IT VOR ORT: KURZVORSTELLUNGEN UND JAHRESRÜCKBLICKE

Wie sieht es eigentlich in den HGF-Zentren und MPG-Einrichtungen aus? Nach der Premiere letztes Jahr wurden wieder insgesamt etwa 20 Kurzvorstellungen verschiedener HGF- und MPG-Einrichtungen über deren Forschung und IT ins „traditionelle“ Programm eingestreut. Schon jetzt sind diese Kurzvorstellungen gefühlte Tradition und erwiesen sich erneut als sehr fruchtbare Grundlage für nachfolgenden Gedankenaustausch – von sehr technischer bis zu völlig untechnischer Art. Besonders war Dirk Viereggs Präsentation des MPI für demografische Forschung in Rostock per Film (<https://nextcloud.init.mpg.de/index.php/s/Pos9rZa9PP4Njdz>). Es zeigte sich einmal mehr, dass allerorten mit gleichartigen Problemen gerungen wird – Personalmangel, Überlastung, immer neue Dokumentationsanforderungen, Bedrohung der Sicherheit, bis hin zu Unverständnis bei der Leitung und am Ende dem Stress, der durch die Anforderungen vor diesem Hintergrund entsteht.

Ein Lichtblick sind jedoch die Ideen und Lösungsansätze, wie diesen Herausforderungen begegnet werden kann. Sie entstehen, weil die administrative IT aus der GV ins Gespräch mit der wissenschaftlichen IT in den Instituten kommt und mit sachkundigen Gästen anderer Wissenschaftsorganisationen, wie z. B. der Fraunhofer Gesellschaft. Es wurden viele nützliche Konturen und Strukturen erkennbarer und es schien, als würden Probleme „Griffe“ bekommen. Daran sollte man dann jetzt anpacken.

## IT MAG DIGITALISIERUNG

IT kann auch Digitalisierung. Es hapert allerdings nur zu oft an der Abstimmung oder Planung, die vorher notwendig ist: Wie sollen Verwaltungsprozesse ablaufen, bei wem ist dieser und bei wem jener Prozess aufgehängt und wer verantwortet ihn? Es wurde positiv wahrgenommen, dass im Bereich der Prozesse in den Administrationen zunehmend Aktivität erkennbar wird. Die IT wäre bereit.

Die Prozesse der Forscher\*innen zum Management ihrer Daten sind trotz erkennbar begonnener Bemühungen noch nicht funktionsfähig in der Praxis angekommen und erfordern wohl noch Geduld bei der IT. Es ist leider noch viel mehr Zuwendung und Aufmerksamkeit bei allen Verantwortlichen nötig. Insbesondere sollten bei allen organisationsweiten Initiativen die IT und Forschungskoordinationen der Zentren und Institute eingebunden werden.

## IMMER WIEDER IT-SICHERHEIT

Ein immerwährendes Thema ist und bleibt die Sicherheit der IT-Systeme im Allgemeinen und der Daten im Besonderen. Berlin, München und Dresden sind unter anderem Orte mit Einrichtungen, die Sicherheitsvorfälle hatten. Dabei traten bei den IT4Science-Days erneut die unterschiedlichen und teils gegensätzlichen Ansätze und Mentalitäten zu Tage. Das Spektrum reichte vom umfänglichen Einsatz von Microsoft-Produkten, proprietären Sicherheitslösungen und der Durchführung von Zertifizierungen bis zu ausschließlichem Einsatz von freier Software und „selbstgestrickten“ Lösungen. Sicher ist trotzdem: Angreifer\*innen können mit genügend Aufwand jedes System übernehmen. Deshalb sind die dezentralen Strukturen der HGF und MPG ein großer Vorteil, denn sie sind resilienter; Zentralisierung sollte stets eine besondere Begründung haben. Im Allgemeinen sollte man ihr skeptisch begegnen.

Die fast täglichen Berichte von erfolgreichen Angriffen aus der ganzen Welt nähren die ewigen Zweifel an der Verlässlichkeit und insbesondere auch der Vertrauenswürdigkeit globaler Computerhersteller und Diensteanbieter wie Microsoft, Google und Apple. Als Produzent von Wissen sind die MPG und HGF bei dieser gesellschaftlichen Frage vermutlich in einer besonderen Verantwortung und Bringschuld.

Für nicht unerhebliche Unruhe sorgte Ende 2022 ein vom BMBF in Auftrag gegebener „Sicherheitsscan“. Das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE hatte ihn durchgeführt. Offene Netzwerkports wurden als große Sicherheitsprobleme dargestellt und die Einrichtungen in ein schlechtes Licht gerückt.

Wie es besser geht, zeigte Tobias Fiebig vom MPI für Informatik in Saarbrücken. Er berichtete über seinen Security-Scan, bei dem er einen etwas über dem Durchschnitt liegenden Stand der Systeme der MPG ermittelt hat. Das ist durchaus bemerkenswert, denn es hatte viele schmerzhaft Sicherheitsvorfälle an Universitäten gegeben und die Befürchtung (oder sogar Unterstellung) stand im Raum, die MPG wäre vielleicht nicht ausreichend „geimpft“, wenn dieser Corona-Vergleich erlaubt ist. Nun zeigt sich, dass die 85 geprüften Institut-ITs in diesem Bereich vergleichsweise gut dastehen.

Das ist auch ohne vorherigen Verdacht, dass es anders sei, bemerkenswert. Bemerkenswert gut. Schließlich handelt es sich um 85 selbstständige Gruppen, die immer wieder, bei Bedarf ausschließlich selbstgesteuert, gut zusammenarbeiten. Gleichzeitig

ist dies kein Ergebnis, das erlauben würde sich auszuruhen. Denn die Sicherheitslage ist weiterhin überaus bedrohlich, wie auch der Sicherheitsvorfall bei der GWDG Ende September gezeigt hat.

## EHRE, WEM EHRE GEBÜHRT

Wie jedes Jahr wurden Kolleg\*innen geehrt, die sich in den vergangenen Jahren um die Gemeinschaft der ITler\*innen in der MPG verdient gemacht haben: Dr. Thomas Otto von der GWDG erhielt den IT-Lebenswerk-Award und Bertin Klein vom MPI für Biologie des Alterns in Köln den IT-Community-Award. Die Laudatio für Dr. Thomas Otto hielt sein langjähriger ehemaliger Kollege Andreas Ißleiber vom MPI für Multidisziplinäre Naturwissenschaften in Göttingen und lobte seine hervorragende, gewissenhafte Arbeit im Hintergrund und seine Bescheidenheit. Knapp 20 DV-Treffen organisierte Dr. Thomas Otto bereits. Dr. Stefan Vollmar vom MPI für Stoffwechselforschung in Köln würdigte die Arbeit von Bertin Klein, der auch langjährig im Organisationskomitee der DV-Treffen aktiv ist und zum Beispiel in letzter Zeit federführend die neuen IT-Leiter\*innen in die IT-Gemeinschaft einführt. Sichtlich bewegten beide die Auszeichnungen entgegen, die wieder liebevoll von Gerry Truschkewitz vom MPI für Dynamik komplexer technischer Systeme in Magdeburg gefertigt worden sind.

## GANZ WEIT DRAUSSEN IM GRÜNEN

Das Konferenzcenter des Max-Delbrück-Centrums in Berlin-Buch war ungewohnt weit draußen an der Stadtgrenze Berlins, bot aber einen sehr guten Rahmen für Kommunikation und Austausch

unter den ca. 170 Teilnehmer\*innen vor Ort. Dazu kamen noch durchschnittlich ca. 80 Online-Teilnehmer\*innen. Die Kolleg\*innen vor Ort, vor allem Jasmin Bonkowski und Michaela Kolbe, hatten alles sehr gut vorbereitet. Hier zeigt sich ein weiterer Unterschied zwischen HGF-Zentren und MPG-Einrichtungen: Die Mehrheit der HGF-Zentren hat auf Grund ihrer Größe ein Konferenzzentrum. In der MPG gibt es dafür nur den Max-Planck-Campus am Faßberg in Göttingen und das Harnack-Haus in Berlin, die mehr als 200 Teilnehmer\*innen gut unterbringen können.

Wie immer waren neben den vielen guten Vorträgen des abwechslungsreichen Programms die Gespräche während der Pausen und am Abend für alle Teilnehmer\*innen von großem Wert. Die Pausen waren diesmal durchgängig mit 30 Minuten eingeplant, was auch genügend Puffer für die manchmal fehlende Zeitdisziplin gab. Die Länge der Tage verlangte Einigen etwas ab. Die Posterpräsentation am Mittwochabend mit gemeinsamer Diskussion und Getränken wurde von einem harten Kern mit Unterhaltung gefüllt bis der letzte Bus fuhr. Für die Verpflegung sorgte The Caternauts Catering und auch dieses Mal wurde das vegane und vegetarische Essen von der großen Mehrheit der Teilnehmer\*innen gelobt. Das spätsommerliche Wetter setzte der sehr gelungenen Veranstaltung dann noch die Krone auf.

Das nächste Treffen wird im kommenden Jahr vom 24. bis 26. September wieder auf dem Faßberg in Göttingen mit den beiden bekannten Gastgebern MPI für Multidisziplinäre Naturwissenschaften und GWDG stattfinden. Informationen zum kennwortgeschützten Download von Präsentationsmaterialien der 2. IT4Science-Days erhalten Sie bei Ihrer lokalen IT. ●

## Kurz & knapp

### Öffnungszeiten des Rechenzentrums um Weihnachten und Neujahr 2023/2024

Das Rechenzentrum der GWDG bleibt an den Tagen vom 24.12. bis zum 26.12.2023 sowie am 31.12.2023 und 01.01.2024 geschlossen. Vom 27.12. bis zum 30.12.2023 ist das Rechenzentrum lediglich von 9:00 bis 17:00 Uhr geöffnet.

Falls Sie sich während der Zeiten, in denen das Rechenzentrum geschlossen ist, an die GWDG wenden möchten, erstellen Sie bitte eine Anfrage über unsere Support-Webseite unter <https://gwdg.de/support> oder schicken eine E-Mail an [support@gwdg.de](mailto:support@gwdg.de). Das dahinter befindliche Ticket-System wird auch während dieser Zeiten von Mitarbeiter\*innen der GWDG regelmäßig überprüft.

Wir bitten alle Nutzer\*innen, sich darauf einzustellen.

Pohl

### Doppelausgabe 1-2/2024 der GWDG-Nachrichten

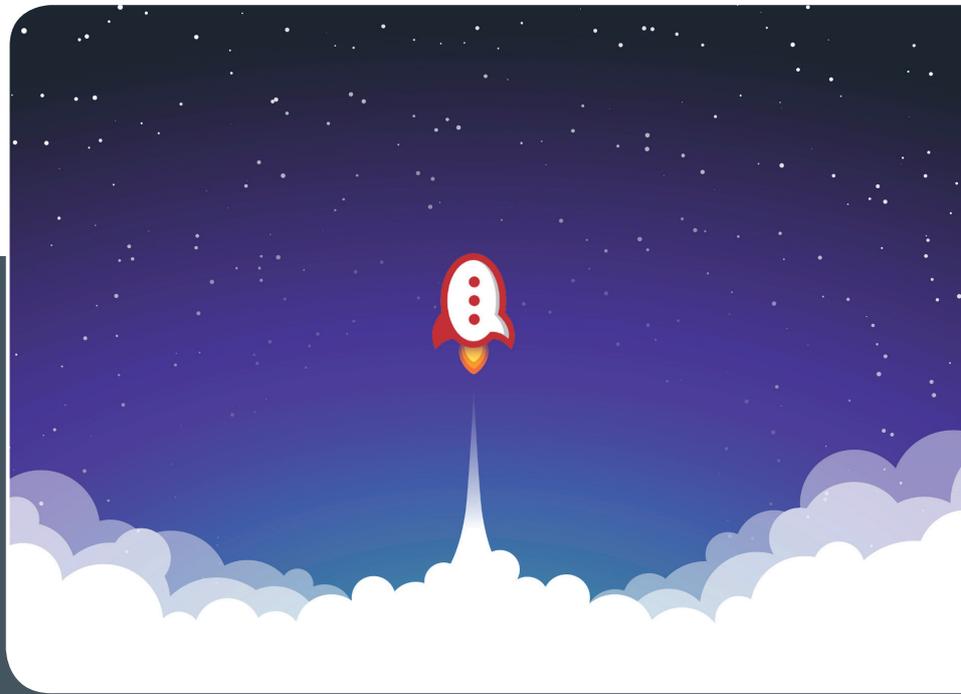
Die nächsten GWDG-Nachrichten erscheinen wie gewohnt als Doppelausgabe 1-2/2024 ca. Mitte Februar 2024.

Otto

### Kursprogramm der GWDG Academy für das erste Halbjahr 2024

Das Kursprogramm der GWDG Academy für das erste Halbjahr 2024 wurde jetzt unter <https://academy.gwdg.de> veröffentlicht. Es enthält eine Vielzahl neuer Kurse, insbesondere aus dem Bereich Scientific Computing bzw. High-Performance Computing. Alle Kurse finden entweder in einem geeigneten Online-Format, das sich in den letzten Jahren bewährt hat, oder als Präsenzkurs statt. Nähere Informationen dazu finden Sie bei den jeweiligen Kursen. Die GWDG Academy präsentiert sich ab sofort in einem komplett neuen Layout und bietet einige interessante neue Features, die Ihnen die Suche nach passenden Kursen sowie die Anmeldung und das Management Ihrer Kurse erleichtern. Wir werden Ihnen die Neuerungen in der GWDG Academy voraussichtlich in der nächsten Ausgabe 1-2/2024 der GWDG-Nachrichten ausführlicher vorstellen. Falls Sie Fragen zur neuen Benutzeroberfläche haben, können Sie sich gerne an [support@gwdg.de](mailto:support@gwdg.de) wenden.

Otto



# Rocket.Chat

KOMMUNIKATION LEICHT GEMACHT!

## Ihre Anforderung

Sie benötigen einen professionellen Chat-Dienst, der eine einfache, persistente Kommunikation mit Kollegen ermöglicht – sowohl in Einzel- als auch in Gruppenunterhaltungen, die komfortabel durchsuchbar sind. Sie wollen Bilder und Dateien mit Kollegen austauschen.

## Unser Angebot

Wir betreiben den Messaging-Dienst „Rocket.Chat“, der es Ihnen ermöglicht, sich in Teams, Gruppen oder auch einzeln auszutauschen. Der Dienst unterstützt zusätzlich Emojis, das Versenden von Dateien, Bildern und Videos sowie die Integration von Benachrichtigungen verschiedener Dienste wie z. B. GitLab. Aufgrund einer breiten Palette von Clients, auch für mobile Geräte, sowie einer übersichtlichen Weboberfläche bieten wir komfortablen Zugriff vom Arbeitsplatz und von unterwegs.

## Ihre Vorteile

- > Einfache Kommunikation im Team
- > Persistente, durchsuchbare Chat-Verläufe
- > Einfaches Teilen von Dateien und Bildern
- > Unterhaltungen mit allen Nutzern, die einen Account bei der GWDG besitzen
- > Integrierte Bots und APIs für die Anbindung von GitLab oder die Einbindung von RSS-Feeds

## Interessiert?

Jeder Nutzer mit einem gültigen Account bei der GWDG und einem aktuellen Webbrowser oder Client kann den Dienst „Rocket.Chat“ nutzen. Für die Benutzung rufen Sie einfach <https://chat.gwdg.de> auf. Nutzer ohne GWDG-Account können einen Account auf <https://www.gwdg.de/registration> registrieren.

# Stellenangebot

Nr. 20231207

**Die GWDG sucht** zum nächstmöglichen Zeitpunkt eine\*in

## Mitarbeiter\*in (m/w/d) für die Leitstelle

mit einer regelmäßigen Wochenarbeitszeit von 39 Stunden und einer täglichen Funktionszeit vorwiegend von 9:00 – 17:00 Uhr. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist bis zur Entgeltgruppe TVöD E 06 vorgesehen. Die Stelle ist nicht zur Teilzeit geeignet und auf zwei Jahre befristet mit der Option auf eine anschließende Entfristung.

### Aufgabenbereiche

- Überwachung und Bearbeitung von elektronischen Störmeldungen im Leitstellensystem (Gemos)
- Bewertung der eingegangenen Informationen und deren Zuordnung
- Weiterleitung von Meldungen, entsprechend Veranlassung nach Vorgabe/Dienstanweisung
- Protokollierung wichtiger Ereignisse, z. B. im Dienstbuch
- Überwachung und Steuerung des Gebäudes, z. B. Steuerung der Vereinzelungsanlagen, Sichtung der Videoüberwachung, Kontrollgänge und Öffnen/ Verschließen des Gebäudes
- Empfang von Personen, teils mit Begleitung zum Zielort, z. B. bei Wartungen der technischen Gebäudeanlagen
- Ausgabe und Verwaltung von Ausweis- und Zutrittskarten (zum Teil mit Kontrolle von Lichtbildausweisen)
- Erfassung und Dokumentation von Anwesenheiten in Sicherheitsbereichen
- Kommunikation mit internen und externen Bereichen/Stellen
- Annahme und Weiterleitung von Meldungen für den Service Desk
- Warenannahme und entsprechende Informationsweitergabe im Haus
- Mitarbeit in der Haustechnik, z. B. Betreuung der Multimediaanlagen
- Verwaltung der Nutzung der Dienstfahrzeuge; Sicherstellen der Einsatzfähigkeit der Dienstfahrzeuge
- Verwaltung von Schlüsseln
- Datenpflege verschiedener Listen
- Vertretung des Empfangs zur Abdeckung einfacher Tätigkeiten
- Teilnahme an Funktionszeiten nach Dienstplan. Langfristig ist eine Ausweitung der Funktionszeiten auf 365 Tage/Jahr, 24 Stunden/Tag und/oder die Teilnahme an der Rufbereitschaft geplant.

### Anforderungen

- Abgeschlossene Berufsausbildung in einem technischen Beruf, vorzugsweise Anlagenmechaniker\*in, Elektroniker\*in für Gebäude und Energie oder Mechatroniker\*in
- Freude an Kommunikation (persönlich und telefonisch) und ein damit verbundenes freundliches und kompetentes Auftreten im Umgang mit Menschen sowie eine schnelle Auffassungsgabe
- Einsatzbereitschaft, Flexibilität und Bereitschaft zur ständigen Weiterbildung
- Allgemeine Computerkenntnisse
- Lösungsorientiertes Handeln sowie die Fähigkeit, in Stresssituationen ruhig und besonnen zu reagieren
- Gute Deutschkenntnisse in Wort und Schrift, Fremdsprachenkenntnisse wie z. B. Englisch sind von Vorteil

### Unser Angebot

- Ein modernes, vielfältiges und inspirierendes Arbeitsumfeld mit großer Nähe zu Wissenschaft und Forschung an der Schnittstelle mehrerer innovativer Technologiesektoren
- Mitarbeit in einem kompetenten und engagierten Team mit kurzen Entscheidungswegen und einem hohen Maß an Eigenverantwortung und Freiheit
- Qualifizierung und Weiterentwicklung Ihrer Fähigkeiten
- Sozialleistungen des öffentlichen Dienstes
- Eine familienfreundliche, offene und lebendige Wissenschaftsstadt mit vielfältigen Freizeit- und Kulturangeboten im Herzen Deutschlands

Die GWDG strebt nach Geschlechtergerechtigkeit und Vielfalt und begrüßt daher Bewerbungen jedes Hintergrunds. Die GWDG ist bemüht, mehr schwerbehinderte Menschen zu beschäftigen. Bewerbungen Schwerbehinderter sind ausdrücklich erwünscht. Es ist zu beachten, dass für die Ausübung von Tätigkeiten in bestimmten Bereichen eine entsprechende körperliche Mobilität Voraussetzung ist.

Haben wir Ihr Interesse geweckt? Dann bitten wir um eine Bewerbung **bis zum 08.01.2024** über unser Online-Formular unter <https://s.gwdg.de/KCXyji>.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

**Herr Stefan Teusch**

E-Mail: [stefan.teusch@gwdg.de](mailto:stefan.teusch@gwdg.de) oder

**Herr Sebastian Pohl**

Tel.: 0551 39-30297

E-Mail: [sebastian.pohl@gwdg.de](mailto:sebastian.pohl@gwdg.de)



## NEUER MITARBEITER GIORGI MAMULASHVILI

Seit dem 1. September 2023 ist Herr Giorgi Mamulashvili als technischer Mitarbeiter in der Arbeitsgruppe „Computing“ (AG C) tätig. Er studierte Physik an der Staatlichen Universität Tiflis in Georgien. Nach seinem Abschluss wechselte er in die IT-Branche. Die wichtigsten Positionen in seiner Laufbahn waren seine Tätigkeiten als Linux-Systemadministrator und als Spezialist für die Entwicklung von Glasfasernetzen. Vor kurzem wurde er vom Linux Professional Institute zertifiziert (LPIC-1 und LPIC-2). Herr Mamulashvili ist per E-Mail unter [giorgi.mamulashvili@gwdg.de](mailto:giorgi.mamulashvili@gwdg.de) zu erreichen.



Kunkel

## NEUER MITARBEITER JAROMIR NEMECEK

Seit dem 15. Oktober 2023 ist Herr Jaromir Nemecek als wissenschaftliche Mitarbeiter in der Arbeitsgruppe „Computing“ (AG C) tätig. Herr Nemecek ist Absolvent der Technischen Universität Brunn (CZ) im Bereich der industriellen Nutzung der Robotik, die er später um einige Semester an der Fakultät für Informatik der Masaryk-Universität erweiterte. 2018 wechselte Herr Nemecek an das International Laser Research Centre (ELI-Beamlines) und war dort an der Implementierung und dem Betrieb von zwei HPC-Systemen für die Simulation der Teilchenphysik beteiligt. In der AG C wird er als Systemadministrator tätig sein. Herr Nemecek ist per E-Mail unter [jaromir.nemecek@dwdg.de](mailto:jaromir.nemecek@dwdg.de) zu erreichen.



Kunkel

## NEUER MITARBEITER MATTHIAS EULERT

Seit dem 1. November 2023 ist Herr Matthias Eulert als wissenschaftlicher Mitarbeiter in der Arbeitsgruppe „Computing“ (AG C) beschäftigt und dort als Trainer für High-Performance Computing (HPC) tätig. Er hat sein Studium der Mathematik an der Georg-August-Universität Göttingen mit dem Diplom abgeschlossen und danach an der Philipps-Universität Marburg und der Universität Kassel als wissenschaftlicher Mitarbeiter gearbeitet. Bei diesen Stationen hat Herr Eulert einerseits wertvolle Erfahrungen in der wissenschaftlichen Lehre bei der Unterrichtung von Studierenden sammeln können und andererseits Kenntnisse im Umgang mit verschiedenen HPC-Systemen erlangt. Er freut sich jetzt darauf, sein Wissen an andere weitergeben zu können und damit einen wesentlichen Beitrag für die HPC-Gemeinschaft vor Ort und im Verbund für Nationales Hochleistungsrechnen zu leisten. Herr Eulert ist per E-Mail unter [matthias.eulert@gwdg.de](mailto:matthias.eulert@gwdg.de) zu erreichen.



Kunkel

INFORMATIONEN:  
support@gwdg.de  
0551 39-30000

Januar bis  
Juni 2024



# Academy

KURS	DOZENT*IN	TERMIN	ANMELDEN BIS	AE
<b>GRUNDLAGEN DER BILDBEARBEITUNG MIT PHOTOSHOP</b>	Töpfer	07.02. – 08.02.2024 9:30 – 16:00 Uhr	31.01.2024	8
<b>VIRTUELLE CLOUD-INFRASTRUKTUREN – KURS FÜR AUSZUBILDENDE</b>	Kopp	13.02. – 15.02.2024 9:00 – 16:00 Uhr	06.02.2024	12
<b>QUICKSTARTING R: EINE ANWENDUNGSORIENTIERTE EINFÜHRUNG IN DAS STATISTIKPAKET R</b>	Cordes	21.02. – 22.02.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr	14.02.2024	8
<b>MYSQL – GRUNDKURS</b>	Groh	28.02.2024 9:30 – 16:00 Uhr	21.02.2024	4
<b>INDESIGN – GRUNDLAGEN</b>	Töpfer	28.02. – 29.02.2024 9:30 – 16:00 Uhr	21.02.2024	8
<b>MYSQL FÜR FORTGESCHRITTENE</b>	Groh	29.02.2024 9:30 – 16:00 Uhr	22.02.2024	4
<b>STATISTIK MIT R FÜR TEILNEHMER*INNEN MIT VORKENNTNISSEN – VON DER ANALYSE ZUM BERICHT</b>	Cordes	06.03. – 07.03.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr	28.02.2024	8
<b>HIGH PERFORMANCE DATA ANALYTICS – PART I</b>	Dr. Ogaja, Nolte, Meisel	06.03. – 07.03.2024 9:30 – 16:00 Uhr	28.02.2024	8
<b>INDESIGN – AUFBAUKURS</b>	Töpfer	12.03. – 13.03.2024 9:30 – 16:00 Uhr	05.03.2024	8

KURS	DOZENT*IN	TERMIN	ANMELDEN BIS	AE
ADMINISTRATION IM ACTIVE DIRECTORY	Quentin, Hast, Kopp	13.03.2024 9:00 – 12:30 und 13:30 – 15:30 Uhr	06.03.2023	4
PRACTICAL COURSE IN HIGH-PERFORMANCE COMPUTING	Prof. Kunkel	02.04. – 05.04.2024 9:00 – 18:00 Uhr 08.04. – 09.04.2024 14:00 – 18:00 Uhr	23.03.2024	20
USING THE GWDG SCIENTIFIC COMPUTE CLUSTER – AN INTRODUCTION	Eulert	10.04.2024 9:30 – 16:00 Uhr	03.04.2024	4
DEEP LEARNING WITH GPU CORES	Sommer, Meisel, Kirchner	11.04.2024 9:30 – 13:00 Uhr	04.04.2024	2
DEEP LEARNING BOOTCAMP: BUILDING AND DEPLOYING AI MODELS	Lewis	15.04. – 16.04.2024 14:30 – 16:30 Uhr	08.04.2024	3
PARALLEL PROGRAMMING WITH MPI	Prof. Haan	16.04. – 17.04.2024 9:15 – 16:00 Uhr	09.04.2024	8
EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS	Cordes	17.04. – 18.04.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr	10.04.2024	8
PERFORMANCE ANALYSIS OF AI AND HPC WORKLOADS	Dr. Ogaja	18.04.2024 9:00 – 12:00 und 13:00 – 16:00 Uhr	11.04.2024	4
GPU PROGRAMMING WITH CUDA – AN INTRODUCTION	Prof. Haan	23.04.2024 9:15 – 16:00 Uhr	16.04.2024	4
STORAGE – KURS FÜR AUSZUBILDENDE	Quentin	24.04. – 25.04.2024 9:00 – 12:30 und 13:30 – 15:30 Uhr	17.04.2024	8
AFFINITY PUBLISHER – SCHNUPPERKURS FÜR EINSTEIGER*INNEN	Töpfer	30.04.2024 10:30 – 12:30 und 13:30 – 15:30 Uhr	23.04.2024	3
SYSTEM, USER AND DEVELOPER PERSPECTIVES ON PARALLEL IO	Höhn, Dr. Krey	07.05.2024 10:00 – 12:00 und 13:00 – 15:00 Uhr	30.04.2024	3
AFFINITY DESIGNER – SCHNUPPERKURS FÜR EINSTEIGER*INNEN	Töpfer	16.05.2024 10:30 – 12:30 und 13:30 – 15:30 Uhr	09.05.2024	3
SNAKEMAKE FOR HPC WORKFLOWS	Paleico	22.05.2024 10:00 – 12:00 und 13:00 – 15:00 Uhr	15.05.2024	3
AGILE SOFTWARE DEVELOPMENT AND DEPLOYMENT USING CONTAINERS	Nordsiek	23.05.2024 10:00 – 12:00 und 13:00 – 15:00 Uhr	16.05.2024	3
DEBUGGING SCIENTIFIC APPLICATIONS – ILLUSTRATION ON OPENFOAM	Höhn, Dr. Ogaja	28.05.2024 10:00 – 12:00 und 13:00 – 15:00 Uhr	21.05.2024	3
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER*INNEN MIT VORKENNTNISSEN	Cordes	29.05. – 30.05.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr	22.05.2024	8

KURS	DOZENT*IN	TERMIN	ANMELDEN BIS	AE
<b>SECURE HPC – PARALLEL COMPUTING WITH HIGHEST SECURITY</b>	Tabougua	31.05.2024 10:00 – 11:30 Uhr	24.05.2024	1
<b>DEEP LEARNING BOOTCAMP: BUILDING AND DEPLOYING AI MODELS</b>	Lewis	03.06. – 04.06.2024 14:30 – 16:30 Uhr	27.05.2024	3
<b>USING JUPYTER NOTEBOOKS ON HPC</b>	Khuzyakhmetov	04.06.2024 9:00 – 12:00 Uhr	28.05.2024	2
<b>GRUNDLAGEN DER BILDBEARBEITUNG MIT AFFINITY PHOTO</b>	Töpfer	05.06. – 06.06.2024 9:30 – 16:00 Uhr	29.05.2024	8
<b>DATA MANAGEMENT CONCEPTS FOR EFFICIENT AND USER-FRIENDLY HPC</b>	Nolte	06.06.2024 9:00 – 12:00 Uhr	30.05.2024	2
<b>BIOINFORMATIC SERVICES AT THE GWDG</b>	Paleico	11.06.2024 9:00 – 12:00 und 13:00 – 16:00 Uhr	04.06.2024	4
<b>HIGH PERFORMANCE DATA ANALYTICS – PART II</b>	Dr. Ogaja, Meisel	12.06. – 13.06.2024 9:30 – 16:00 Uhr	05.06.2024	8
<b>USING THE GWDG SCIENTIFIC COMPUTE CLUSTER – AN INTRODUCTION</b>	Eulert	17.06.2024 9:30 – 16:00 Uhr	10.06.2024	4
<b>ANSYS ON CLUSTER AND POST-PROCESSING OF SIMULATION RESULTS</b>	Höhn, Dr. Kanning	18.06.2024 9:00 – 12:00 und 13:00 – 16:00 Uhr	11.06.2024	8
<b>QUICKSTARTING R: EINE ANWENDUNGSORIENTIERTE EINFÜHRUNG IN DAS STATISTIKPAKET R</b>	Cordes	19.06. – 20.06.2024 9:00 – 12:00 und 13:00 – 15:30 Uhr	12.06.2024	8

#### Teilnehmerkreis

Das Angebot der GWDG Academy richtet sich an die Beschäftigten aller Einrichtungen der Universität Göttingen, der Max-Planck-Gesellschaft sowie aus wissenschaftlichen Einrichtungen, die zum erweiterten Kreis der Nutzer\*innen der GWDG gehören. Studierende am Göttingen Campus zählen ebenfalls hierzu. Für manche Kurse werden spezielle Kenntnisse vorausgesetzt, die in den jeweiligen Kursbeschreibungen genannt werden.

#### Anmeldung

Für die Anmeldung zu einem Kurs müssen Sie sich zunächst mit Ihrem Benutzernamen und Passwort in der GWDG Academy (<https://academy.gwdg.de>) einloggen. Wenn Sie zum Kreis der berechtigten Nutzer\*innen der GWDG gehören, erhalten Sie anschließend automatisch Zugang zu unserem Kursprogramm. Sollten Sie noch keinen Account besitzen, können Sie sich unter <https://id.academiccloud.de> registrieren und müssen ggf. auf Anfrage für die Anmeldung zu unseren Kursen freigeschaltet werden. Bei Online-Kursen kann das Anmeldeverfahren abweichen. Genauere Informationen dazu finden Sie in der jeweiligen Kursbeschreibung. Einige Online-Angebote stehen Ihnen jederzeit und ohne Anmeldung zur Verfügung.

#### Absage

Absagen können bis zu sieben Tagen vor Kursbeginn erfolgen. Bei kurzfristigeren Absagen werden allerdings die für den Kurs angesetzten Arbeitseinheiten (AE) vom AE-Kontingent der jeweiligen Einrichtung abgezogen.

#### Kursorte

Die Kurse finden entweder in einem geeigneten Online-Format oder als Präsenzkurs statt. Nähere Informationen dazu finden Sie bei den jeweiligen Kursen. Auf Wunsch und bei ausreichendem Interesse führen wir auch Kurse vor Ort in einem Institut durch, sofern dort ein geeigneter Raum mit entsprechender Ausstattung zur Verfügung gestellt wird.

#### Kosten bzw. Gebühren

Die Academy-Kurse sind – wie die meisten anderen Leistungen der GWDG – in das interne Kosten- und Leistungsrechnungssystem der GWDG einbezogen. Die den Kursen zugrundeliegenden AE werden vom AE-Kontingent der jeweiligen Einrichtung abgezogen. Für alle Einrichtungen der Universität Göttingen und der Max-Planck-Gesellschaft sowie die meisten der wissenschaftlichen Einrichtungen, die zum erweiterten Kreis der Nutzer\*innen der GWDG gehören, erfolgt keine Abrechnung in EUR. Dies gilt auch für die Studierenden am Göttingen Campus.

#### Kontakt und Information

Wenn Sie Fragen zum aktuellen Academy-Kursangebot, zur Kursplanung oder Wünsche nach weiteren Kursthemen haben, schicken Sie bitte eine E-Mail an [support@gwdg.de](mailto:support@gwdg.de). Falls bei einer ausreichend großen Gruppe Interesse besteht, könnten u. U. auch Kurse angeboten werden, die nicht im aktuellen Kursprogramm enthalten sind.



Gesellschaft für wissenschaftliche  
Datenverarbeitung mbH Göttingen