

# GWDG NACHRICHTEN 09-10|23

Sicherheitsvorfall

Academic Cloud Hub

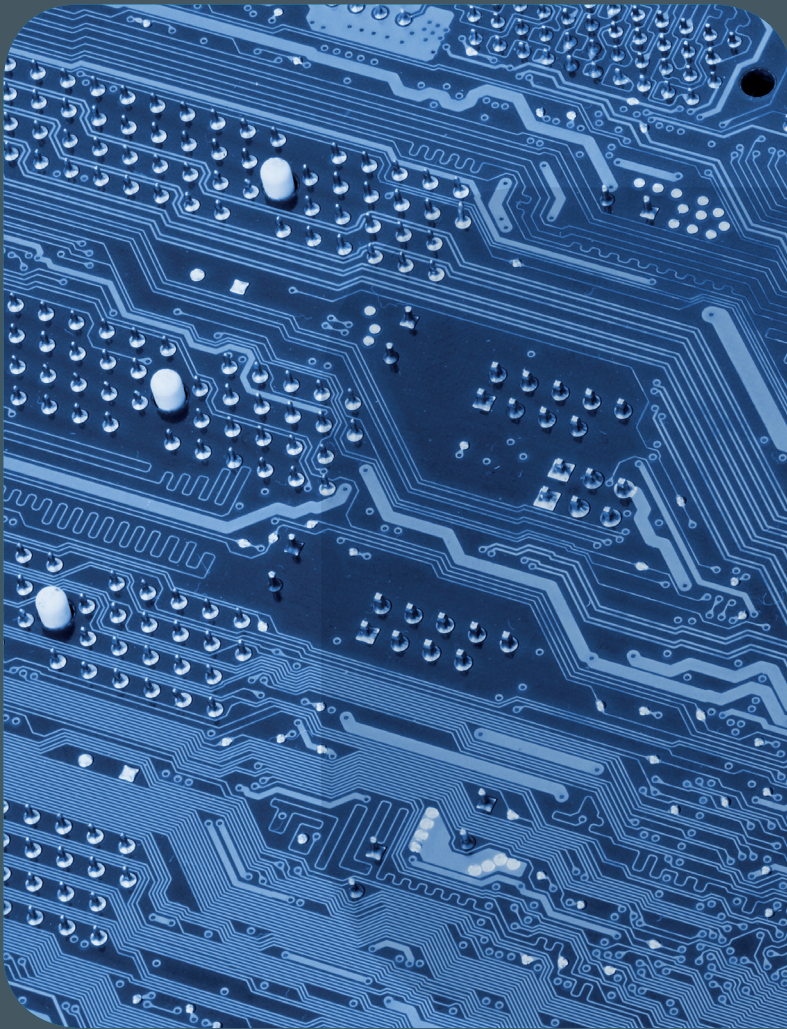
GöHPC Admin Tea

SynForest

Metadata Extraction  
with Web Crawlers

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG





## GWDG NACHRICHTEN

# 09-10|23 Inhalt

- .....
- 4 **Sicherheitsvorfall vom 28. September 2023**
  - 5 **Academic Cloud Hub – ein flexibles soziales Netzwerk**
  - 8 **Von Admins für Admins – der GöHPC Admin Tea**
  - 9 **Kurz & knapp**
  - 10 **SynForest – Synthetic Generation of LiDAR Data in Forests**
  - 14 **Metadata Extraction with Web Crawlers in Education**
  - 17 **Stellenangebot**
  - 18 **Personalia**
  - 22 **Academy**

### Impressum

.....

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686  
46. Jahrgang  
Ausgabe 9-10/2023

**Erscheinungsweise:**  
10 Ausgaben pro Jahr

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

**Auflage:**  
550

**Fotos:**

© cherezoff - stock.adobe.com (1)  
© chagin - Fotolia.com (13)  
© nito - Fotolia.com (17)  
© pterwort - Fotolia.com (21)  
© Robert Kneschke - Fotolia.com (22)  
© Nataliya Kalabrina - Fotolia.com (23)  
© MPI-NAT-Medienservice (3, 21 unten)  
© GWDG (2, 18, 19, 20, 21 oben)

**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,*

*im Editorial habe ich schon häufiger zur allgemeinen Bedrohungslage und der Bedeutung von Informationssicherheit berichtet. Wir hatten dies zum Anlass genommen, unser Sicherheitsmanagement immer wieder auszubauen und Anfang des Jahres auch die zugehörige ISO-27001-Zertifizierung zu erlangen, was uns bei der Bewältigung des Sicherheitsvorfalls im September geholfen hat. Vor allem hat aber der schnelle und unermüdliche Einsatz des gesamten Teams der GWDG und unserer IT-Partner\*innen sowie der IT-Verantwortlichen in den Einrichtungen unserer Kund\*innen geholfen, dass die Auswirkungen im Vergleich zu ähnlichen Vorfällen bei anderen Einrichtungen im Wissenschaftssystem überschaubar blieben. Mit dem kurzfristigen Neusetzen von Passwörtern konnte ein etwaiges Missbrauchsrisiko minimiert werden. Dennoch bedauern wir den Vorfall und eventuelle Unannehmlichkeiten bei unseren Kund\*innen. Dies ist ein Ansporn, auch weiter an einem Ausbau des Sicherheitsmanagements zu arbeiten.*

**Ramin Yahyapour**

*GWDG – IT in der Wissenschaft*

# Sicherheitsvorfall vom 28. September 2023

**Text und Kontakt:**  
Presse- und Öffentlichkeitsarbeit  
der GWDG  
oeffentlichkeitsarbeit@gwdg.de

Die Zahl der Cyber-Angriffe in Deutschland liegt auf hohem Niveau und sind eine wachsende Bedrohung für Wirtschaft, Politik und Wissenschaft. Nach Angaben des Bundeskriminalamts (BKA) wurden im Jahr 2022 rund 140.000 Fälle registriert. Auch die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) hat am 28. September 2023 durch interne Sicherheitssysteme einen Sicherheitsvorfall festgestellt. Mit Hilfe des internen Informationssicherheitsmanagements und des zugehörigen Monitorings konnte ein Angriff auf die Infrastruktur in einem frühen Stadium identifiziert und weitere Auswirkungen minimiert werden.

## MANAGEMENT DER SITUATION

Zur Aufklärung und Analyse des Vorfalls hat die GWDG alle notwendigen Ressourcen bereitgestellt: Innerhalb weniger Stunden wurde mit Cyber-Security-Expert\*innen und externen Sicherheitsexpert\*innen ein Einsatz-Team gebildet und die Arbeit aufgenommen. Es wurden kurzfristig Gegenmaßnahmen zum weiteren Schutz der Infrastruktur vorgenommen. Zudem hat die GWDG die entsprechenden Datenschutzbehörden sowie die Gesellschafter\*innen und Kund\*innen umgehend informiert.

Eine wichtige Maßnahme zur Bewältigung der Situation war die schnelle Änderung aller Passwörter und die Sperrung von Accounts, bei denen die Passwörter nicht geändert wurden. Die von der GWDG angebotenen Dienste waren weitgehend verfügbar. Einzelne Dienste, die im Rahmen einer Vorsichtsmaßnahme vorübergehend deaktiviert worden sind, wurden bereits wieder verfügbar gemacht bzw. werden so bald wie möglich wieder bereitgestellt.

Die forensische Untersuchung wurde am 11. Oktober 2023 abgeschlossen. Nach der umfassenden Prüfung von Diensten, Systemen und Daten konnten die externen Expert\*innen Entwarnung geben.

## WEITERE ERHÖHUNG DER IT-SICHERHEIT

Zusammen mit den Sicherheitsexpert\*innen wurden verschiedene Maßnahmen ergriffen, um die Sicherheit der relevanten IT-Systeme über das vorhandene hohe Niveau hinaus zusätzlich abzusichern und das Informationssicherheits-Managementsystem der GWDG weiter auszubauen. Zusätzlich wird die Überwachung der Infrastruktur vorsorglich weiterhin verstärkt.

Wir möchten unseren Kund\*innen und ihren Beschäftigten an dieser Stelle herzlich für ihre Unterstützung im Umgang mit dem Sicherheitsvorfall danken.

Für eventuell entstandene Unannehmlichkeiten entschuldigen wir uns noch einmal ausdrücklich und danken Ihnen für Ihr Verständnis. Ein besonderer Dank geht auch an unsere IT-Partner\*innen sowie die IT-Verantwortlichen in den Einrichtungen unserer Kund\*innen für die konstruktive Zusammenarbeit und Unterstützung. ●

## Security Incident on September 28, 2023

On September 28, 2023, a security incident was detected on GWDG systems by internal security systems. With the help of the internal information security management and the associated monitoring, an attack on the infrastructure was identified at an early stage and further effects were minimized. Together with external security experts, countermeasures were taken at short notice to further protect the infrastructure. Various measures were taken to additionally secure the security of the relevant IT systems beyond the existing high level and to further expand the GWDG's information security management system. One important measure to deal with the situation was the quick change of all passwords and the blocking of accounts where passwords had not been changed. The services offered by the GWDG were largely available. The forensic investigation was completed on October 11, 2023. After the comprehensive examination of services, systems and data, the external experts\* were able to give the all-clear. Nevertheless, the monitoring of the infrastructure will continue to be intensified.

# Academic Cloud Hub – ein flexibles soziales Netzwerk

## Text und Kontakt:

Tobias Herbst  
tobias.herbst@gwdg.de  
0551 39-30296

Ralph Krimmel  
ralph.krimmel@gwdg.de  
0551 39-30257

Seit 2018 bietet die Academic Cloud ein stetig wachsendes Angebot an Diensten, welche den Austausch von Daten und die virtuelle Zusammenarbeit erleichtern sollen. Die neueste Ergänzung ist HumHub als Alternative zu Facebook, um Lehrende und Lernende an den Hochschulen in Niedersachsen zusammenzubringen und Communities zu finden, die sich zu einem speziellen Thema oder Projekt austauschen.

## HINTERGRUND

Der Bedarf nach digitalen Kommunikations-, Organisations- und Vernetzungsmöglichkeiten ist nach wie vor hoch und wächst in Zeiten von projekt- und ortsübergreifenden Kollaborationen und auch Homeoffice stetig. Oft wird für diese Zwecke auf gängige soziale Netzwerke zurückgegriffen, die vernetzte Kommunikation und Gruppenzusammenarbeit ermöglichen. Diese sind hinsichtlich des Datenschutzes aber mindestens zweifelhaft und durch die Finanzierung über Werbung vor allem darauf ausgerichtet, die Verweildauer der Nutzer\*innen auf der Plattform zu maximieren. Eine alternative Möglichkeit bieten daher soziale Plattformen, die auf eigenen Servern gehostet werden können. Mastodon gilt z. B. als Alternative zu Twitter und HumHub als Alternative zu Facebook.

## GWDG STARTET ACADEMIC CLOUD HUB

In den Projekten Connex, Co<sup>3</sup>Learn, Futur.A und Souver@n entstand die Idee HumHub einzusetzen, um Lehrende und Lernende aus den Hochschulen in Niedersachsen zusammenzubringen. Die GWDG hat daher eine HumHub-Instanz in der Academic Cloud eingerichtet und damit steht dieser Dienst allen Nutzer\*innen zur Verfügung. Die neue Plattform soll die Lücke zwischen Chats und Microblogging (z. B. Mastodon oder Twitter/X) und den auf ein konkretes Projekt bezogenen Diensten wie GitLab oder OpenProject füllen. Hier kann man eine Community finden und sich zu einem Thema austauschen, ohne eine Längenbegrenzung der Texte beachten oder direkt in ein Projekt einsteigen zu müssen.

## WAS IST HUMHUB?

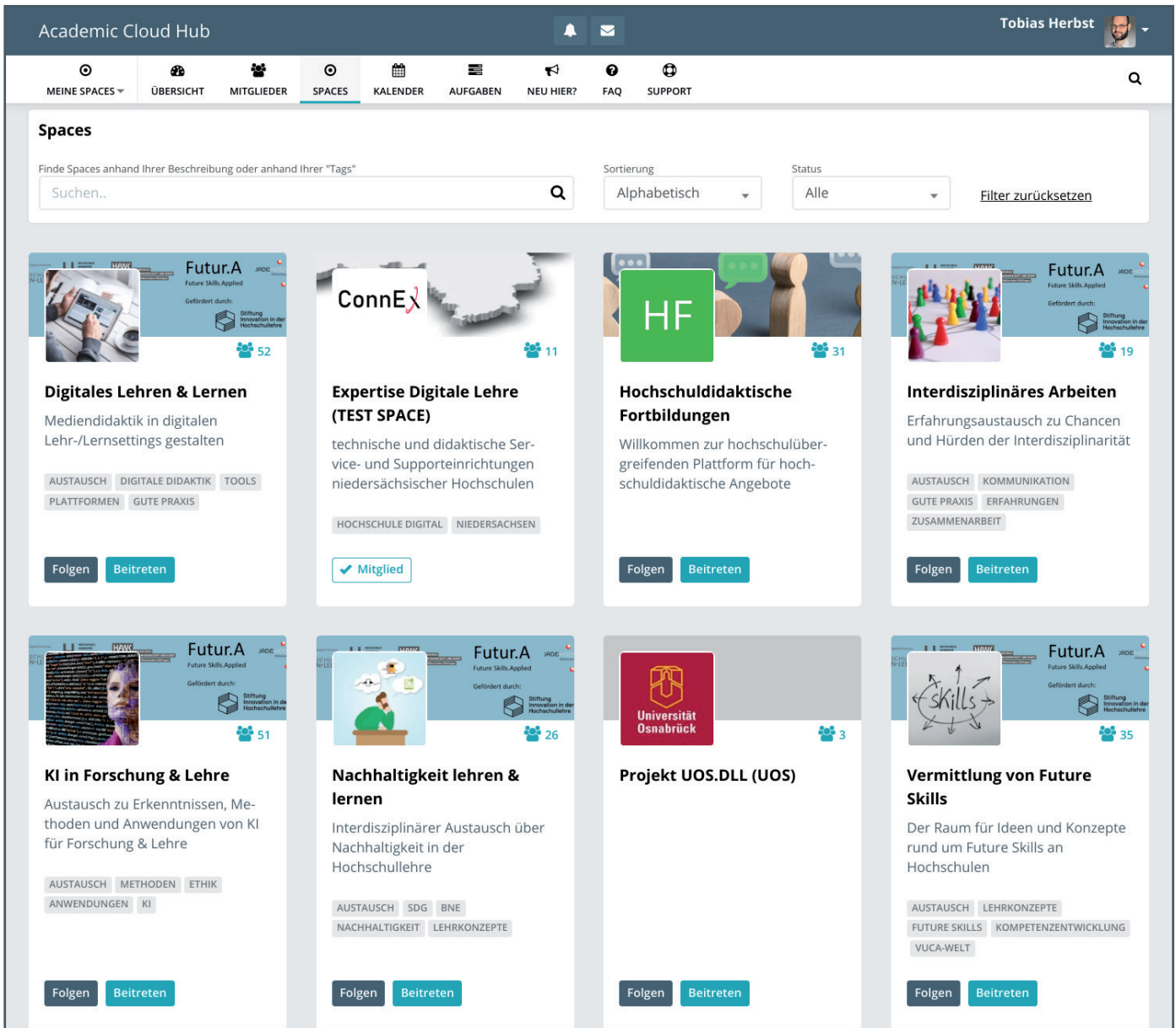
HumHub [1] ist eine Open-Source-Software zur Einrichtung eines sozialen Netzwerks, das aber auch für die interne Unternehmenskommunikation und -zusammenarbeit gut geeignet ist und als Basis für ein Social Intranet angepasst werden kann. Es wurde ursprünglich von einer deutschen Webagentur entwickelt und ist mittlerweile ein eigenständiges Unternehmen. HumHub ist an die Bedienung von Facebook angelehnt, bietet jedoch flexiblere Einsatzmöglichkeiten, ist durch Module erweiterbar und so für den jeweiligen Einsatz individuell anpassbar. Durch das Hosten auf eigenen Servern ermöglicht es zudem einen datenschutzkonformen Einsatz.

## SPACES, PROFILE UND MODULES – FEATURES VON HUMHUB

In HumHub findet Kommunikation in sogenannten „Spaces“ statt (siehe Abbildung 1). „Spaces“ können einem Projekt, Gruppen oder Themen gewidmet sein, zu denen sich die Nutzer\*innen austauschen wollen. Jeder „Space“ kann mit individuellen Rechten und Regeln ausgestattet werden und Nutzer\*innen können „Spaces“ beitreten oder dazu eingeladen werden. „Spaces“ können entweder privat, öffentlich für angemeldete Nutzer\*innen oder

### Features im Überblick

- **Dashboard**  
Schneller Überblick über die wichtigsten Informationen
- **Social Tools**  
Folgen, Liken, Erwähnungen und Vernetzung mit anderen
- **Nutzerprofil**  
Die digitale Visitenkarte für jede\*n Nutzer\*in inklusive Anlegen von Portfolios, persönliche Vernetzungen oder Teilen von Statusmeldungen auf der eigenen Wall
- **Gruppen**  
Zur Organisation in Niederlassungen, Abteilungen o. ä.
- **Directory**  
Übersichtlicher Einstieg zum Finden von Spaces, Nutzer\*innen und Gruppen
- **Suche**  
Zum Auffinden von Nutzer\*innen, Diskussionen und Dateien
- **Freier Zugriff**  
Inhalte sind auch mit nicht registrierten Benutzer\*innen teilbar
- **Notifications**  
Benachrichtigungen über Themen, die Sie als Nutzer\*in direkt betreffen
- **Activity Stream**  
Zum Verfolgen von Diskussionen und Fortschritten
- **Sprachanpassung**  
HumHub wurde mittlerweile in über 18 Sprachen internationalisiert



1\_Liste der öffentlichen Spaces im Academic Cloud Hub

komplett öffentlich (auch für nicht angemeldete Besucher\*innen) sein. So kann Gruppen- und Teamkommunikation ähnlich wie bei Facebook stattfinden – mit dem Unterschied, dass HumHub-Spaces durch Tools und Module um zusätzlich benötigte Funktionalitäten erweitert werden können. Im HumHub-Marketplace gibt es beispielsweise Module für Kalender, Aufgabenplanung, Galerien, Wikis, gemeinsame Dateibearbeitung, Dateiablagen und Datenuploads. Diese Module lassen sich für jeden „Space“ einzeln an- oder ausschalten und konfigurieren. Alternativ können auch eigene Erweiterungen programmiert und integriert werden.

## NUTZUNG DES ACADEMIC CLOUD HUB

Durch die Einbindung in die Academic Cloud ist der Zugriff auf HumHub sehr einfach. Studierende und Mitarbeiter\*innen benötigen für ihre Anmeldung lediglich eine gültige Benutzerkennung (Benutzername und Passwort) ihrer niedersächsischen Hochschule. Nach der Anmeldung in HumHub entweder indirekt über das AC-Portal [2] und dem entsprechenden Icon oder direkt über [3] werden Sie automatisch im „Willkommen“-Space hinzugefügt.

Dort gibt es in der Menüleiste die Einträge „Neu hier?“ und „FAQ“. „Neu hier?“ bietet eine kurze Einführung in die Funktionen und gibt Hilfe beim Einstieg in die Nutzung des Academic Cloud Hub. „FAQ“ beantwortet die wichtigsten Fragen, die sich ergeben haben. Auf der Übersichtsseite (siehe Abbildung 2) werden die Beiträge aus allen Spaces, denen Sie beigetreten sind, aufgelistet.

## Academic Cloud Hub: Selfhosted Macroblogging

Since 2018, the Academic Cloud has offered a steadily growing range of services designed to facilitate data sharing and virtual collaboration. The latest addition is HumHub as an alternative to Facebook to bring together teachers and learners at universities in Lower Saxony and find communities to share ideas on a specific topic or project. This new service fills a gap between microblogging and project-oriented services. Users can write longer posts or comments, create a community and coordinate with them.

The screenshot shows the Academic Cloud Hub interface. At the top, there is a navigation bar with the title 'Academic Cloud Hub' and a user profile for 'Tobias Herbst'. Below the navigation bar are several tabs: 'MEINE SPACES', 'ÜBERSICHT', 'MITGLIEDER', 'SPACES', 'KALENDER', 'AUFGABEN', 'NEU HIER?', 'FAQ', and 'SUPPORT'. The main content area is divided into two columns. The left column features two user profiles: Katja Hoffmann and Thessa Hackbarth. Katja's post is titled '#servicemonday' and discusses creating posts. Thessa's post is titled '#community' and discusses Spaces. The right column contains three sections: 'Erste Schritte' (First Steps), 'Letzte Aktivitäten' (Recent Activities), and 'Anstehende Veranstaltungen' (Upcoming Events). The 'Erste Schritte' section lists three guides: 'Anleitung: Überblick', 'Anleitung: Spaces', and 'Anleitung: Benutzerprofil'. The 'Letzte Aktivitäten' section shows a list of recent events, including 'Willkommen im Academic Cloud Hub' and 'Anleitung: Spaces'. The 'Anstehende Veranstaltungen' section shows an event titled 'Virtuelle Führung durch den ACHub' on 30.08.2023. At the bottom of the right column, there is a footer: 'Förderung · Powered by HumHub'.

2\_Hauptseite des „Willkommen im Academic Cloud Hub“-Spaces

## TECHNISCHER HINTERGRUND, LIZENZIERUNG UND INSTALLATION

HumHub basiert auf dem PHP-Framework Yii, die Grundlage ist ein traditioneller LAMP-Stack (Linux, Apache, MySQL und PHP). Es kann sowohl als Software as a Service (SaaS) vom Hersteller gehostet werden als auch selbst auf eigenen Webservern installiert werden, wodurch HumHub zu einer datenschutzfreundlichen Version eines sozialen Netzwerkes wird. In der Professional Edition sind zusätzlich für Unternehmen relevante Funktionen (z. B. für E-Mail-Whitelists) enthalten. Außerdem gibt es kostenpflichtige Tools für Umfragen, zur Theme-Gestaltung oder für Massen-Benachrichtigungen. Die Lizenzierung erfolgt über die „HumHub Commercial License“, was es ermöglicht, Erweiterungen zu programmieren, die nicht wie in der Free Version an die Community weitergegeben werden müssen.

## ÜBER DIE ACADEMIC CLOUD

Die Academic Cloud [2] steht seit 2018 als Kollaborationsplattform für Studierende und Mitarbeiter\*innen niedersächsischer Hochschulen zur Verfügung und stellt zahlreiche Dienste zum Datenaustausch sowie zur virtuellen Zusammenarbeit in Bildung und Forschung bereit. Sie wird von der GWVG gehostet. Das Projekt wurde in Zusammenarbeit mit dem Landesarbeitskreis Niedersachsen für Informationstechnik / Hochschulrechenzentren (LANIT/HRZ; [4]) realisiert und wird durch das Niedersächsische Ministerium für Wissenschaft und Kultur gefördert.

## LINKS

- [1] <https://www.humhub.com>
- [2] <https://academiccloud.de>
- [3] <https://hub.academiccloud.de>
- [4] <https://www.lanit-hrz.de>

# Von Admins für Admins – der GöHPC Admin Tea

## Text und Kontakt:

Christoph Hottenroth  
christoph.hottenroth@gwdg.de

Der Verein „Nationales Hochleistungsrechnen“ (NHR) ist ein Zusammenschluss von Hochleistungsrechenzentren verschiedenster Universitäten, Institute und Hochschulen. Im Rahmen dieses Vereins kam auch der Wunsch nach einem professionellem Austausch der Personen, die am Betrieb und der Administration beteiligt sind, auf. Hierfür wurde eine monatliche, virtuelle Veranstaltung geschaffen, die es ermöglichen soll, über die verschiedensten Aspekte der HPC-Administration zu informieren und sich auszutauschen. Der Name wurde in Anlehnung an die in Göttingen bereits etablierte auf HPC-User fokussierte Veranstaltung (GöHPCoffee) gewählt: GöHPC Admin Tea, kurz GöHAT.

## AUSGANGSLAGE UND ZIELSETZUNG

Zusammen mit der Universität Göttingen ist die GWDG eines von bundesweit neun Zentren für Nationales Hochleistungsrechnen (NHR-Zentren). Speziell im Fokus steht der Ausbau der Kompetenzen in den Bereichen der Lebenswissenschaften, Erdsystemwissenschaften, Strömungsmechanik, Künstlichen Intelligenz und Big Data und der digitalen Geisteswissenschaften. Hierbei bildet die GWDG gemeinsam mit dem Zuse-Institut Berlin (ZIB) und ihrem gemeinsamen Fachberatungs-Netzwerk den NHR-Nord.

Aktuell betreibt die GWDG das NHR-System „Emmy“ und die GPU-Erweiterung „Grete“, die beide bereits in früheren Ausgaben der GWDG-Nachrichten vorgestellt wurden. Das NHR-Schwester-System „Liese“ wird beim ZIB betrieben.

Ziel des GöHAT ist es nun, die Administratoren der verschiedenen NHR-Mitgliedszentren zusammenzubringen, um einen Austausch von Know-how und somit die Entwicklung im HPC-Bereich zu fördern.

## ABLAUF EINES GÖHAT

Die monatlichen Treffen des GöHAT finden jeweils am ersten Donnerstag um 9:30 Uhr statt und werden über die von der GWDG betriebene BigBlueButton-Plattform abgehalten. Der Ablauf eines GöHAT-Meetings beginnt mit einer Begrüßung und

einer kurzen Einführung. Anschließend wird ein Vortrag von etwa 30 bis 45 Minuten Dauer zu einem oder mehreren Themen mit HPC-Bezug präsentiert. Diese Vorträge decken eine breite Palette von Themen ab und wurden bereits von Kolleg\*innen der GWDG sowie des ZIB vorgestellt.

Nach dem Vortrag folgt eine interaktive Frage- und Antwortrunde, die in der Regel zunächst auf das präsentierte Thema eingeht, aber auch in verwandte Themenbereiche übergehen kann. Diese Diskussionen tragen dazu bei, die Expertise der Teilnehmer\*innen weiter zu vertiefen.

Je nach Vortragendem und Zuhörerschaft werden die Präsentationen und Fragerunden in Deutsch oder Englisch abgehalten.

Die GöHAT-Veranstaltung steht nicht nur HPC-Administratoren offen, sondern auch allen anderen, die Interesse am Thema Hochleistungsrechnen haben. Selbst wenn Sie nicht aktiv an der Diskussion teilnehmen möchten, sind Sie herzlich eingeladen, als Zuhörer\*in an den Präsentationen und Diskussionen teilzunehmen.

Interessierte Personen können sich unter [1] in die zugehörige Mailingliste eintragen, um Einladungen und weitere Informationen zu kommenden Veranstaltungen zu erhalten.

## ÜBERBLICK BISHERIGER THEMEN

Der erste GöHAT startete am 2. März 2023 mit einem Vortrag von Freja Nordsiek, Mitarbeiterin der Arbeitsgruppe „Computing“ (AG C) der GWDG, mit dem Thema „HPC Setup for SIF Containers



1\_Logo des GöHAT

## GöHAT for Admins

The GöHPC Admin Tea is a monthly gathering of HPC Administrators. Every Month on the first Thursday at 9:30 am, they meet via BigBlueButton to exchange experiences, developments or new practical use cases for established workflows. These meetings are intended mainly for Administrators working in HPC, but on the other hand, everyone interested is welcome to participate, even passively by just attending and listening to the Presentation and the discussion afterward.



## A barebones sandbox

Start with (almost) nothing ...

```
$ mkdir -p empty/etc
$ touch empty/etc/{passwd,group}
$ singularity build empty.sif empty
```

... and use what singularity adds:

```
$ singularity build --sandbox sandbox empty.sif

$ ls -a sandbox/
.  ..  bin  dev  environment  etc  .exec  home  proc  root  .run
.shell  singularity  .singularity.d  sys  .test  tmp  var
```

2\_Auszug aus dem GÖHAT-Vortrag „Small experiments with Nix and NixOS“

with Apptainer“. Azat Khuzyakhmetov, ebenfalls Mitarbeiter der AG C, gab daraufhin mit „Building software and providing modules with Spack“ einen Überblick über den flexiblen Paketmanager Spack und seinen Einsatz bei der GWDG, sowie eine Einführung zum Bauen, zur Verfügungstellung und zum Verwalten von Modulen damit.

Thomas Poschwatta vom Zuse-Institut Berlin (ZIB) stellte den Teilnehmer\*innen den Paketmanager Nix sowie die darauf aufbauende Linux-Distribution NixOS vor und zeigte unter anderem die Ergebnisse seines Experiments, einen SLURM-Server auf Basis dieser Distribution zu erstellen und zu betreiben (siehe Abbildung 2).

Mit „Stateful PXE-Boot using Warewulf“ führte uns Hendrik Nolte von der AG C in das OS-Provisionierungstool ein, welches primär auf den GWDG-eigenen Clustern zum Einsatz kommt.

Zuletzt gab uns Tobias Watermann vom ZIB eine Übersicht über die verschiedenen Möglichkeiten, einen HPC-Cluster stromsparender zu betreiben.

### AUSBLICK

Um den Informationsaustausch weiter zu fördern, sind in der Zukunft zusätzliche Kommunikationskanäle geplant. Dazu gehören ein eigener Rocket.Chat-Channel sowie ein Mastodon-Account. Eine Webseite mit einem Überblick über die nächsten Veranstaltungen steht bereits zur Verfügung und kann unter [2] besucht werden.

Zusätzlich wird aktiv nach spannenden Vortragsthemen mit HPC-Bezug gesucht. Wenn Sie ein solches Thema vor einem interessierten Publikum präsentieren möchten, können Sie sich gerne direkt per E-Mail an [christoph.hottenroth@gwdg.de](mailto:christoph.hottenroth@gwdg.de) wenden.

### LINKS

- [1] <https://listserv.gwdg.de/mailman/listinfo/goehat>
- [2] <https://gwdg.de/hpc/events/goehat/> 

## Kurz & knapp

### IT-Security-Awareness-Days vom 07.11. – 17.11.2023

Mehrere deutsche Hochschulen, u. a. auch die Universität Göttingen / GWDG, veranstalten gemeinsam vom 07.11. – 17.11.2023 zum sechsten Mal die IT-Security-Awareness-Days als hochschulübergreifende öffentliche Online-Veranstaltungsreihe mit spannenden Vorträgen rund um das Thema Informationssicherheit. Die Vortragsthemen reichen von Deepfakes, ChatGPT & Co. bis hin zum richtigen Umgang mit Passwörtern

und richten sich vorrangig an Nutzer\*innen ohne besondere Vorkenntnisse. Zudem werden technische Vorträge angeboten, die auch für Expert\*innen und Administrator\*innen von Interesse sein könnten. Die Veranstaltung ist öffentlich, eine Anmeldung ist nicht erforderlich. Eine Zusammenfassung der Veranstaltungstermine steht unter dem URL <https://s.gwdg.de/9NyyDI> bereit. Informationen zu den einzelnen Vorträgen finden Sie unter dem URL <https://www.tu-braunschweig.de/ciso/it-sad/it-sad-wintersemester-2023-24>.

Otto

# SynForest – Synthetic Generation of LiDAR Data in Forests

## Text and Contact:

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

As part of the BMBF founded project FORESTCARE we have developed SynForest, an application that generates high-quality point cloud datasets of synthetic forests. For this purpose, we aimed to produce realistic patterns by combining forest growth simulation tools such as ForestFactory with tree shape modeling based on real data from existing databases such as PyTreeDB. Additionally, the scanning method, trajectory, and sensor properties are simulated in Helios++ to approximate real imaging techniques. SynForest is able to utilize the power of our HPC cluster by distributing the task among nodes and achieves efficiency through parallelism for arbitrarily large datasets. Our results show that the synthetic datasets closely resemble real-world data and therefore SynForest can be regarded as a practical and cost-effective alternative for the study of large-scale forest environments and tree structures, with many potential applications in forestry, deep learning and data analysis.

## INTRODUCTION

Point cloud data is commonly used to describe the 3D structure and shape of real-world environments intricately and accurately. However, acquiring high-quality data poses many challenges that can impede the progress of large projects. Imaging is hindered by the cost of obtaining and operating equipment such as drones and LiDAR sensors. Point clouds often need to be preprocessed which requires substantial effort and expertise. In the forestry domain, segmenting individual trees is typically required for subsequent feature extraction and analysis, adding to the complexity of this task. As raw data is a prerequisite for most tasks, data acquisition can become a major bottleneck if not handled properly.

For this purpose, we developed an application that generates high-quality point clouds of synthetically produced forests quickly and cost-efficiently (see Figure 1). Our goal is to enable researchers to produce as much data as desired with customizable forest configurations that resemble target forest environments. By simulating forest growth in ForestFactory, building an artificial 3D forest environment using real tree models and simulating a LiDAR sensor in Helios++, we aimed to produce synthetic datasets that resemble real data as closely as possible. With SynForest, we hope that researchers can preprocess and analyze the data, prepare workflows, and pretrain deep learning models much more efficiently than acquiring real data.

## METHOD

The synthetic forest production is divided into three main steps: Map, Scene, and Simulation. A JSON configuration file is

used to determine the settings of each step (see Figure 2).

### Map

During the map step, a relatively simple map of the forest is produced, containing the coordinates, heights, and diameters, and species of the trees in the forest.

An initial forest map is produced using ForestFactory. ForestFactory uses climate records to simulate the growth of trees over a period of several years and incorporates interactions between the trees. The result is a tabular data file containing the position, height, species, and diameter of each tree in the forest. ForestFactory can produce forests with several tree species and variable

## SynForest – Synthetische Erzeugung von LiDAR-Daten in Wäldern

Im Rahmen des BMBF-geförderten Projektes FORESTCARE entwickelten wir Synforest, eine Anwendung, die hochwertige Punktwolkendatensätze synthetischer Wälder generiert. Es werden realistische Daten mit einer Kombination aus dem Simulationstool ForestFactory, der Datenbank PyTreeDB sowie des Scanprogramms Helios++, welches reale Bildgebungstechniken annähert, erzeugt. SynForest kann auf dem HPC-Cluster durch das Parallelisieren der einzelnen Schritte beliebig große Datensätze simulieren. Mithilfe dieses Programms können in den Bereichen Forstwissenschaft und Machine Learning Methoden der Datenanalyse getestet und verbessert werden.

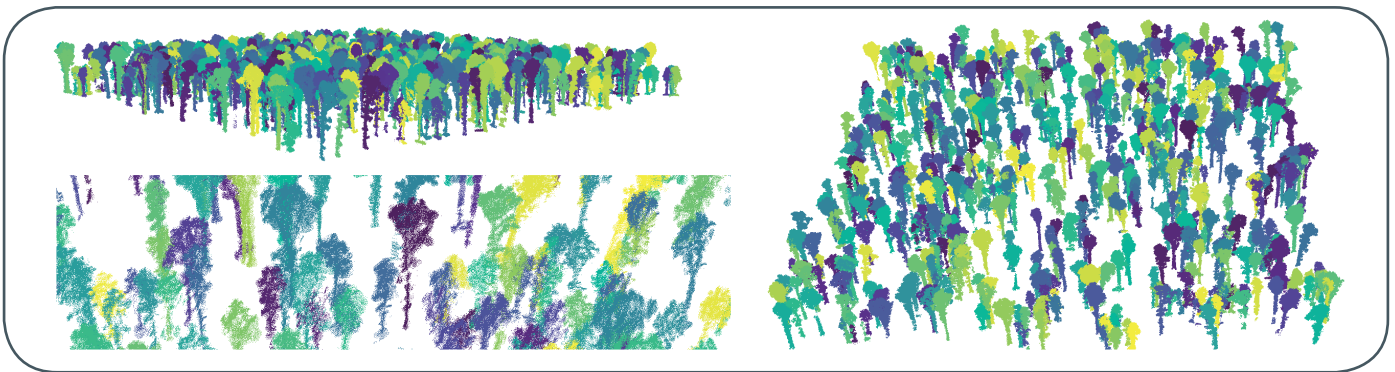


Figure 1: Example of a forest point cloud produced with SynForest. Individual trees are colored randomly (Top-left: Side view, Bottom-left: Zoomed in, Right: Full view)

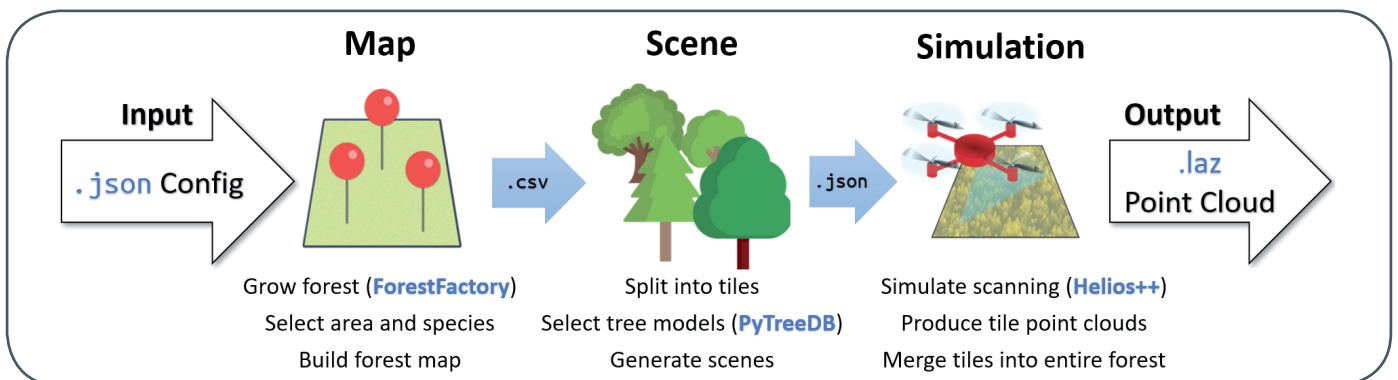


Figure 2: Overview of steps involved in SynForest's workflow

TREE ID	X	Y	DIAMETER	HEIGHT	SPECIES
1	3.2266	39.0349	0.4137	34.4867	FagSyl
2	45.1323	103.0373	0.5894	44.1235	PicAbi
3	84.5816	76.3333	0.4564	35.9486	FagSyl
4	155.5193	134.1912	1.2078	37.7912	Que-Rub
5	53.9892	33.8612	0.2593	26.0365	Que-Pet

Table 1: Example of data rows produced in map step

densities. With appropriate climate records, it is possible to simulate forest growth in any environment.

The initial map produced by ForestFactory is then preprocessed to select the target area and species as stated in SynForest's configuration file. The result of this step is a CSV file consisting of the positional information for each individual tree in the forest map (see Table 1). At this point, the global structure of the forest is determined, but the individual tree shapes are unknown, except for general information, e.g., the tree's height and diameter.

### Scene

In the scene step, the 3D forest environment is fully determined, including the precise shapes and models of the trees. The forest is split into several smaller tiles for the purpose of parallelization, and a JSON scene file is produced for each tile, describing the 3D models and orientations for each tree.

Taking the forest map produced in the previous step as input, the target area is tiled according to the settings in the configuration file. The tiles may overlap with each other by a specified amount, to prevent edge case artifacts on borders.

Within each tile, three-dimensional tree models are placed at their corresponding positions. To find a suitable model for each

tree, its species, height, and diameter are considered as described in the forest map. SynForest then searches for the closest match in a database of real tree point clouds, e.g., PyTreeDB. Trees can be chosen with or without leaves and the minimum scanning quality can be set in the configuration file. The models may also be scaled or randomly rotated if desired. The final 3D model of each tree is then placed in the tile's scene file.

Once the scene files are generated, an XML file is produced for each scene for the purpose of simulation in Helios++.

### Simulation

In the third and final step, simulation, the individual tile scene files are loaded in parallel on multiple HPC nodes, and by simulating the imaging method, high-quality point clouds are generated for each tile.

Using the flight path and the sensor's parameters as determined in SynForest's configuration file, the corresponding XML survey files are created for simulation with Helios++. A job array is submitted using Slurm to run Helios++ in parallel and produce a point cloud for each tile.

When all simulations are completed, the point clouds are merged into a single point cloud, excluding the overlapping points to prevent multiple occurrences on the tile borders. The points have many features, including intensity, return number, and tree ID. Along with the final point cloud, a complementary CSV file is produced that describes the individual trees and their features, which can be used to select desirable points in the point cloud.

### USAGE

The installation guide is available in SynForest's repository [1].

Once SynForest is installed, it can be run simply with:

```
python run.py config/example_ULS.json
```

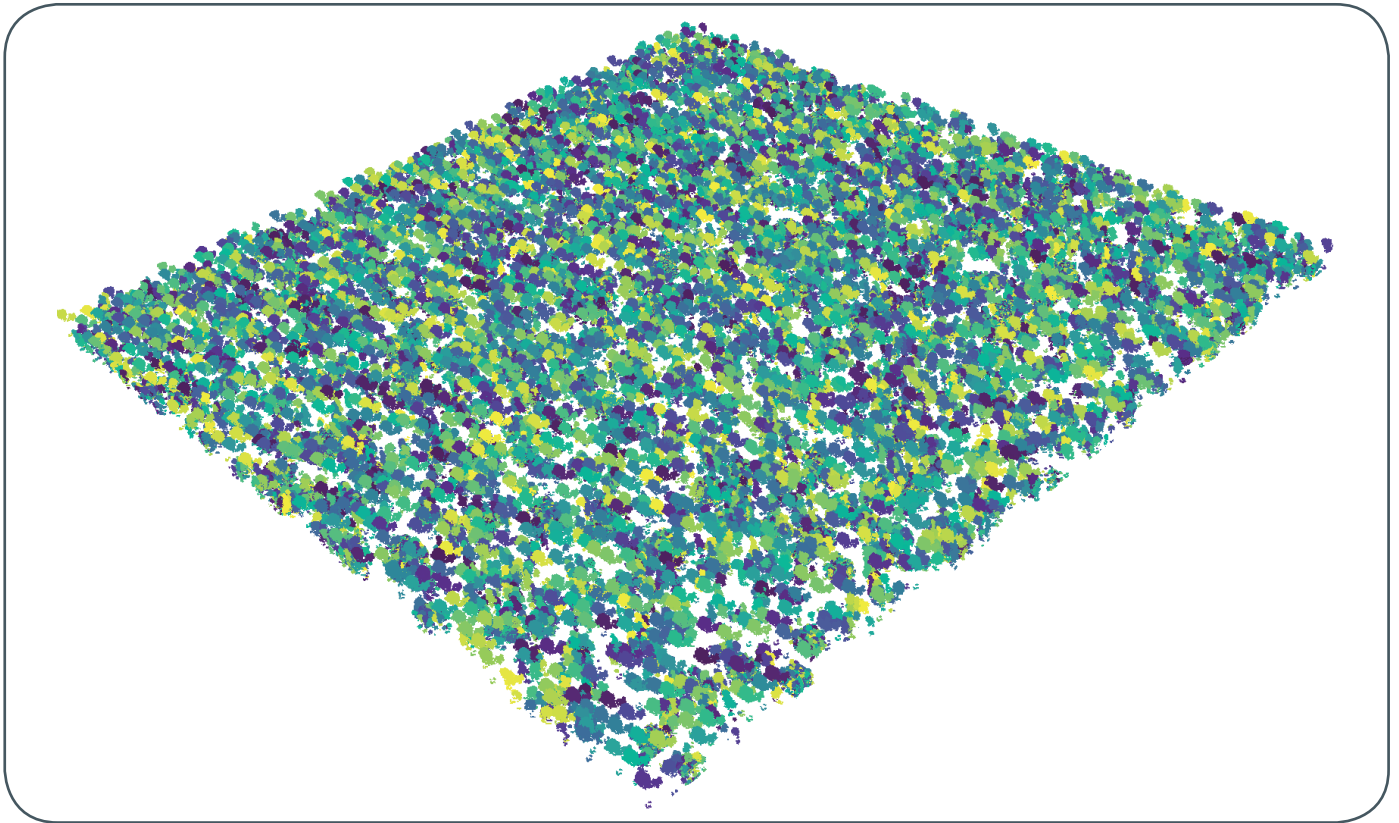


Figure 3: Example of a synthetically generated forest. This forest plot is 1,000 m x 1,000 m, consisting of 37,915 trees and approximately 110 million points (1.7 GB)

PARAMETER	DESCRIPTION
<i>side</i>	Side length of the square map in meters
<i>species</i>	Identifier(s) of the tree species, e.g. <i>QuePet</i>
<i>tile_w, tile_h</i>	Width and height of each tile in meters
<i>tile_overlap</i>	The overlap between adjacent tiles in meters
<i>platform</i>	ID, speed, and altitude of simulated platform (e.g., drone or plane) used for the survey
<i>scanner</i>	ID and parameters of simulated lidar sensor (e.g., Zenmuse L1)

Table 2: The most essential parameters that users can define to generate LiDAR data according to their requirements

The script loads the configuration file in the argument and submits a number of slurm jobs. Depending on the forest size, configuration, and available nodes, the task may take several minutes or hours. Once the final job is completed, a full-size point cloud (see Figure 3) and its complementary CSV file will be stored in the *results* folder.

## CONFIGURATION

Many parameters can be adjusted in SynForest by specifying them in the configuration file. A detailed explanation of all configurable parameters can be found at [2].

A few notable parameters are shown in Table 2.

## APPLICATIONS

As a data synthesis tool, SynForest has endless applications. Owing to its simple usage and efficiency, it is possible to produce many large datasets rapidly. These synthetic datasets can be used as an alternative to real data for various purposes.

In machine learning and deep learning, SynForest can be used to generate datasets for developing pipelines and pretraining models for a variety of forestry tasks. Since each point in the point cloud has a tree ID, its species and other tree features can be extrapolated using the complementary CSV file. Therefore, the datasets can be used to train and evaluate segmentation and classification models.

The data is also suitable for forest analysis. Since accuracy and realism is a priority in every step of SynForest’s workflow, the resulting files are comparable to real forest data. For analysis tasks where sufficient data cannot be obtained directly, it is possible to produce similar forests using only the target’s climate records as an approximation to real data. Furthermore, the parameters in the configuration file can be altered to produce custom forests with different species and densities. To obtain a synthetic forest with a specific shape, one can simply produce a larger square with SynForest and then select the trees within the desired shape.

SynForest’s large data output and job parallelism can also be taken advantage of for system benchmarking and enhancement.

SynForest is available at [3].

## ACKNOWLEDGEMENTS

- [1] <https://gitlab-ce.gwdg.de/hpc-team-public/synforest>
- [2] <https://gitlab-ce.gwdg.de/hpc-team-public/synforest/-/blob/main/docs/CONFIG.md>
- [3] <https://gitlab-ce.gwdg.de/hpc-team-public/synforest>
- [4] ForestFactory: <https://git.ufz.de/angermue/forestfactory>
- [5] PyTreeDB: <https://pytreedb.geog.uni-heidelberg.de/>
- [6] Helios++: <https://github.com/3dgeo-heidelberg/helios>
- [7] Janika Schäfer et al.: “Applications of synthetic airborne laser scanning data of forests”, ForestSat Conference 2022. ●



# 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 GWDG-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.

# Metadata Extraction with Web Crawlers in Education

## Text and Contact:

Andrés Granados

andres-felipe.granados-cruz@gwdg.de

Data availability from websites is being improved each time with the quality of the metadata. Metadata defines the specific content of a website and web crawlers are a powerful tool to extract this data and the main content of a website. Here it is discussed how the metadata can be extracted with web crawlers, their advantages and pitfalls.

## METADATA IN EDUCATIONAL WEBSITES

Information on the internet is divided by websites where each piece of information is represented by an URL. We don't have a knowledge base if we only account for the data contained in websites. But if we are able to identify each type of content and link them to other entities (URLs) we can get a knowledge base with which to classify and index new data. The metadata is represented in the following structures and standards to make data from websites more machine readable:

- **RDF:** Resource Description Framework is a graph data model that allows to represent contextual data about resources on the internet. RDF provides a grammar for its syntax which can be written in different formats such as Turtle, TriG, N-Triples, N-Quads and JSON-LD [1].
- **JSON-LD:** JSON for Linking Data is a serialization format used to represent an instance of a RDF data model [1].
- **Microdata:** Consists of nested groups (items) of name-value pairs (properties) that are added to documents. Microdata is used, for example, by Google Search as support of the structured data [2].
- **DOM:** Document Object model is the representation of the data and structure of a website through nodes and objects. The DOM can be accessed and modified using JavaScript.
- **WP-JSON:** Wordpress is a CMS (Content Management System) which provides the WP-JSON REST API. This API provides REST endpoints representing the posts, pages, taxonomies, and other data types. Other applications can communicate with the Wordpress REST API by means of JSON data from and to those endpoints to query, modify and create content on a site [3].

Metadata standards are specified by open communities such as Schema.org [4] or Google Search Engine [5]. They offer, for example, validators for the metadata markups:

- Schema Markup Validator [6]
- Rich Results Test of the Google Structured Data Markups [7]

Metadata for educational institutions is very valuable, it allows the institution to offer specific educational content that learners and teachers can reach easily and efficiently. In this way information

can be spread on the internet allowing us to track its availability and limitations. The metadata fields that are more useful are:

- **Learning Resource Type:** Is the predominant type of resource. Examples: presentation, video, material.
- **EducationalLevel:** Level or stage of a learning content which is suitable for a determined educational context. Examples: primary, secondary education.
- **Audience:** Corresponds to the educational audience role or the target group for which the resource is intended to be taught. Examples: student, parent, teacher.
- **License:** Contains an id (URI) that links to the specific license such as CC, GNU, Apache, etc.
- **Description:** Two or three sentences explaining the content of the webpage. Ideally it should contain the main topic, intended user or audience.
- **Keywords:** To describe the content.
- **Publisher:** The responsible entity for publishing the content. It contains an URI which identifies the publisher.
- **inLanguage:** Language of the content.

Some of the metadata fields mentioned above should belong to a controlled vocabulary that defines and limits the values that the fields should have such that the web scrapers can find them and then the editors can index and curate them.

## WEB CRAWLERS

Web crawlers and web scrapers are internet bots which visit and index one or many URLs from a website. By using Scrapy,

### Metadatenextraktion mit Webcrawlern im Bildungsbereich

Die Verfügbarkeit von Daten aus Webseiten wird mit der Qualität der Metadaten immer weiter verbessert. Metadaten definieren den spezifischen Inhalt einer Webseite und Webcrawler sind ein leistungsfähiges Werkzeug, um diese Daten und den wesentlichen Inhalt einer Webseite zu extrahieren. Im Folgenden wird der Ansatz, wie Metadaten mit Webcrawlern extrahiert werden können, erörtert sowie dessen Vor- und Nachteile diskutiert.

a Python based framework that allows to define “spiders” (programs that can extract, catalog and index data from an URL) those repetitive steps are tailored to get access to tables, forms, links and metadata. And due to the specific representation and location of their data and metadata in each website, those steps determine how to access the data structures, HTML templates, and the methods the webpage uses to load the data.

A Spider defines how an URL will be scraped following this basic steps:

- Define a root URLs or set of entries to be visited
- Then the crawler fetches other webpages (Seeds) and makes a list of them (Horizon)
- Define the callbacks for each URL
- Then the response retrieves the objects
- Parsing the data and create a tree/structure of the HTML data
- Pull the data out of the parse tree (by using for example BeautifulSoup [8])
- Those actions can be monitored and it offers a powerful way to look for changes in the content of the websites.

There are many levels in which a developer or a general user can implement a web crawler task. From a crawler framework like Scrapy that allows a developer to define the extraction, transformation and loading tasks (ETL process) of the data until the ready-to-use visual crawlers that are available in the market and that allow any user to define, schedule and export its scrapped resulting data.

## DEVELOP A WEB CRAWLER WITH SCRAPY

Scrapy is a python based framework for web scraping. The workflow starts from the spiders that send requests to the engine and the scheduler of all their requests, then it is yielded to the downloader that fetches the data from the internet. Then it is followed by the data parsing which is managed again by the spiders, then the extracted data is mapped, cleansed and persisted in the items pipeline.

But additional tools such as headless browsers are required to successfully get the data in case the webpages uses Single-Page-Applications architecture that reloads only some components of a webpage showing the requested data to the client.

The core method to locate the data in the DOM of a website is by using XPath selectors which are path expressions to select and extract nodes in a XML file.

## VISUAL CRAWLERS

A very useful tool to avoid writing specific scripts for web crawling usually devoted to developers, is to allow any user to select the sources of data, and design a detailed workflow for all the ETL process easily. By using visual crawlers users are allowed to build specific workflows for their desired data to extract, transform and load the desired data into files or using cloud storage services.

Visual crawlers allow us to simulate the steps needed to extract data as a human can do. It allows us to fill login forms, search prompts, switch drop down menus, scroll in a list web page, capture data from web objects such as text, links, images or the entire HTML code. Some websites render Javascript routines before their content data is completely loaded and some

visual crawlers are capable of loading those websites in headless browsers. Also some solutions offer APIs and webhooks (callback scripts intended to complete other tasks after the extractions by the crawler) of the web crawling tasks to extend their functionality, and also integrate visual crawlers with other platforms and frameworks. The available visual crawlers and APIs have different pricing strategies, most of them based on the number of concurrent processes that are performed in the cloud and on cloud storage.

## ADDITIONAL CONSIDERATIONS

Another important security and efficiency variables for developing or using a crawler are:

- Concurrency
- Control the website rate limiting with IP rotation or Proxy settings to avoid being blocked by the websites that detect non-human interactions trying to access to the website content
- Implement User-agents
- Configure proxy servers: The number of requests can be controlled to avoid a single IP address being identified and blocked by the website.

It is important to consider the use of these techniques to avoid being blocked in the context of scraping a website efficiently when looking for the target information and not with the purpose to generate an attack or any related damage to a website or organization.

## LOCATE OR GENERATE SITEMAPS

The web crawling process should start with a detailed review of all the relevant URLs to be crawled. Most of the time this list of URLs is not available and the map of the website can be very difficult to build from scratch. Usually an efficient approach is to find out the pattern of all the potential URLs using a list of URLs, from a search component in the target website. Then a script can loop over all the pages, access each of them and then extract the target data based on a HTML template. There are two main steps to improve the website exploration before the ETL process having a sitemap, by using:

- A sitemap finder: To scan the entire website and look for an XML sitemap. For example with tools like Sitemap Test [9].
- A sitemap generator: In the worst case there is no sitemap file then it can be created with sitemap generator [10].

## MEASURING THE QUALITY OF KEYWORDS IN METADATA

The quality of the keywords metadata from educational sources can be quantified by using similarity metrics basically performed by two approaches:

- Lexical similarity which scores how two texts are related based on the single characters, words or n-grams (n sequence of characters or words); and by
- Semantic similarity that quantifies how two texts are related by their meaning.

The lexical similarity techniques can lead to very accurate

results if it's not required to analyze the text meaning. It looks for similarities by means of the vector representation of each word. It is faster but sensitive to spelling, punctuation and synonyms.

The semantic similarity is more complex to implement than the lexical similarity. It is not sensitive to synonyms or the order of the words used. Then semantic similarity can find relations between two texts given their contexts and concepts even if they use different words.

Usually the keywords used in metadata are general terms (not verbal conjugation) or nouns with few variations then both similarity approaches can be applied. Keywords and other metadata fields like Disciplines, Learning Resource Type or Educational Level are only one or very few number of words, the text similarity can be simplified and the meaning of the texts in this case is not a restriction.

Two applications of the two approaches for measuring similarity are:

- Wordnet path similarity
- Word vector similarity

By using the Wordnet in the Wordnet [11] and NLTK [12] packages in Python, and using the Open German WordNet (odnet v1.3) it is possible to measure how close the generated metadata is relative to a controlled vocabulary. For example, the distance of the path between two terms can be estimated using the Path similarity, which results in 1.0 when the synsets are the same, and 0.0 when the terms are not related or if the synset doesn't belong to the German WordNet (or belongs to other WordNet like English, for example).

A document (sentences) can be represented as vectors too, then two vectors can be compared, by using for example, the cosine angle and then compute the angle between them (how similar they are). This metric doesn't account for the size of the vector, but it accounts for its direction. This means that two vectors:

- can have exactly the same representation, similarity =  $s = 1$
- would not be related, the words in one document are not related to the words in the other document,  $s = 0$
- or they can have totally opposite representations, their direction is totally opposed,  $s = -1$
- or they can have any combination of those three cases (similarity should be one value between  $-1$  and  $1$ )

## CONCLUSIONS

Web crawling allows the educational institutions and learners to increase the efficiency of finding specific learning or teaching

content in every website resource. In order to find specific information in a network, some internet organizations have developed standards which every webpage should conform to. Those standards define the wide variety of metadata types and categories and the allowed values that each metadata should have. Following the metadata schema specifications for each type of content and providing a sitemap is a key tool to allow the website to have machine-readable content and therefore making its content available and easily accessible for everyone on the internet.

Web crawling can be performed by tailored scripts using for example Scrapy a framework for scraping and web crawling, or by solutions already built to make web crawling a task that everyone can perform.

Keywords and other single-term metadata can be compared to a reference vocabulary by using similarity. The analysis of the similarity scores depends on the intrinsic meaning (semantic analysis) or in their vector representation (lexical).

## REFERENCES

- [1] JSON-LD  
<https://www.w3.org/TR/json-ld/#relationship-to-rdf>
- [2] HTML standard – Microdata  
<https://html.spec.whatwg.org/multipage/microdata.html#microdata>
- [3] WordPress REST API handbook  
<https://developer.wordpress.org/rest-api/>
- [4] Schema.org  
<https://schema.org/>
- [5] Google Search Engine – Structured search data  
<https://developers.google.com/search/docs/appearance/structured-data/intro-structured-data>
- [6] Schema Markup Validator  
<https://validator.schema.org/>
- [7] Rich Results Test of the Google Structured Data Markups  
<https://search.google.com/test/rich-results>
- [8] BeautifulSoup  
<https://beautiful-soup-4.readthedocs.io/en/latest/>
- [9] Sitemap finder tool  
<https://seositecheckup.com/tools/sitemap-test>
- [10] Sitemap generator  
<https://github.com/lgraubner/sitemap-generator>
- [11] Interface for Wordnets  
<https://wn.readthedocs.io/en/latest/index.html>
- [12] Natural Language Toolkit  
<https://www.nltk.org/>



# Stellenangebot

Nr. 20231025

**Die GWDG sucht** ab sofort zur Verstärkung des Storage-Teams der Arbeitsgruppe „IT-Infrastruktur“ (AG I) eine\*n

## Storage Engineer (m/w/d)

mit einer regelmäßigen Wochenarbeitszeit von 39 Stunden. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist je nach Qualifikation bis zur Entgeltgruppe TVöD E 13 vorgesehen. Die Stelle ist zur Teilzeit geeignet und zunächst auf zwei Jahre befristet mit der Option auf Verlängerung. Die GWDG strebt eine langfristige Zusammenarbeit an.

Die GWDG betreibt in ihren modernen Rechenzentren mehrere Massenspeichenumgebungen mit einer Bruttokapazität von insgesamt mehr als 60 PB auf HDDs/SSDs und mehr als 110 PB auf Band für Archivierungs- und Backupzwecke. Diese Massenspeichenumgebungen bilden die Basis nicht nur für die üblichen IT-Basisdienste, sondern insbesondere auch für moderne Cloud-Native-Architekturen, Kubernetes-Cluster und innovative Applikationen aus der Spitzenforschung mit speziellen Anforderungen. Viele Großprojekte und die Nähe zu Wissenschaft und Forschung eröffnen dabei spannende Möglichkeiten zur kreativen Fortentwicklung dieser Umgebungen nahe am State of the Art.

### Aufgabenbereiche

- Mitarbeit im Storage-Team der GWDG zur Sicherstellung eines stabilen und qualitativ hochwertigen Betriebs der unterschiedlichen Massenspeicher- und Backup-Umgebungen der GWDG
- Konzeptionierung und Weiterentwicklung dieser IT-Infrastrukturen mit einem hohen Maß an Selbstständigkeit und Eigenverantwortung
- Konsequente Optimierung durch Automatisierung von Prozessen und Einsatz von Methoden aus dem DevOps-Umfeld („Infrastructure as Code“)
- Beratung externer und interner Kund\*innen und eigenverantwortliches Management von entsprechenden Projekten
- Mitwirkung beim 2nd/3rd-Level-Support

### Anforderungen

- Abgeschlossenes technisches Hochschulstudium (Master/Diplom) oder eine abgeschlossene Berufsausbildung mit adäquater beruflicher Erfahrung
- Programmier-/Skriptingkenntnisse (z. B. Python oder vergleichbar) oder Erfahrung mit Infrastructure-as-Code-Frameworks (z. B. Terraform, Puppet oder Ansible)
- Fundierte Kenntnisse in der Systemadministration unter Linux

- Ein grundlegendes Verständnis von modernen Infrastrukturkonzepten ist wünschenswert (SDN, SDS, Objectstorage/S3, Container, Virtualisierung etc.)
- Erfahrungen mit einem oder mehreren der folgenden Produkte/Systeme sind von Vorteil: TSM/ISP, Bareos, Ceph, MinIO, NetApp FAS/AFF/E-Series, Quantum StorNext, IBM Spectrum Scale / GPFS
- Einsatzbereitschaft, Flexibilität und Bereitschaft zur ständigen Weiterbildung
- Ausgeprägte Kund\*innen- und Serviceorientierung

### Unser Angebot

- Flexible Arbeitszeiten und die Möglichkeit zu mobilem Arbeiten
- Ein modernes, vielfältiges und außergewöhnliches 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
- Unterstützung bei der Qualifizierung und Weiterentwicklung Ihrer Fähigkeiten in hochinteressanten und stark nachgefragten Zukunftsfeldern
- 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. Haben wir Ihr Interesse geweckt? Dann bitten wir um eine Bewerbung **bis zum 22.11.2023** über unser Online-Formular unter <https://s.gwdg.de/tyUVPo>.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

**Herr Piotr Kasprzak**

Tel.: 0551 39-30101

E-Mail: [piotr.kasprzak@gwdg.de](mailto:piotr.kasprzak@gwdg.de)



## 25-JÄHRIGES DIENSTJUBILÄUM DR. CHRISTIAN BOEHME

Herr Dr. Christian Boehme schloss sein Studium der Chemie an der Universität Marburg mit einer Promotion im Bereich der Computerchemie ab. Nach einem Auslandsaufenthalt in Strasbourg und einem Postdoc an der Universität Bochum kam er 2003 als wissenschaftlicher Mitarbeiter zur GWDG. Zu einer seiner ersten Aufgaben gehörte die Einführung eines neuen Ressourcen-Managements und Monitorings für die HPC-Systeme der GWDG. Später führte er die verschiedenen HPC-Cluster-Inseln und das HPC-Hosting der GWDG in einem gemeinsamen Cluster-Management zusammen. Bei der GWDG war Herr Dr. Boehme für die beiden Drittmittelprojekte OptiNum-Grid und Cloud4E verantwortlich, die sich mit der Erforschung von Plattformen für Grid- und Cloud-gestütztes Engineering befassten, sowie für das Projekt ProfiT-HPC, in dem eine Lösung für User-zentriertes Monitoring von HPC-Workloads entwickelt wurde. Zu seinen Aufgaben gehörte außerdem die Organisation der Euro-Par 2019 in Göttingen. Ab 2013 leitete er das HPC-Team der GWDG. In dieser Rolle war er maßgeblich an den Anträgen zum Supercomputer Emmy und zum Göttinger NHR-Zentrum NHR-NORD@Göttingen beteiligt. Zudem leitete er GWDG-seitig die Planung des modularen Data Centers (MDC), in dem Emmy aufgebaut wurde. Seit 2022 ist er stellvertretender Leiter der neugegründeten Arbeitsgruppe „Computing“ (AG C) und arbeitet intensiv an neuen, innovativen Compute-Diensten für die Nutzer\*innen der GWDG. Wir bedanken uns bei Herrn Dr. Boehme für seine geleistete Arbeit und gratulieren ihm herzlich zu seinem 25-jährigen Dienstjubiläum. Vor allem aber freuen wir uns auf das, was wir in Zukunft noch gemeinsam erreichen werden.

Kunkel



## ABSCHIED VON DR. VANESSA END

Frau Dr. Vanessa End hat vom 1. Januar 2013 bis zum 31. Oktober 2016 zunächst als Doktorandin in der Arbeitsgruppe „eScience“ (AG E) in Kooperation mit der T-Systems Solutions for Research GmbH am BMBF-Projekt GASPI zur Forschung an asynchroner und einseitiger Kommunikations-API für HPC gearbeitet. Das Thema ihrer am 1. Juni 2017 erfolgreich abgeschlossenen Promotion lautete „On Collective Communication and Notified Read in the Global Address Space Programming Interface (GASPI)“. Vom 1. Januar 2017 bis zum 31. August 2023 war Frau Dr. End als wissenschaftliche Mitarbeiterin u. a. im Projekt ProfiT-HPC und in verschiedenen Funktionen im Bereich NHR/HLRN tätig, ab 1. Dezember 2021 in der neuen Arbeitsgruppe „Computing“ (AG C), in die das HPC-Team überführt wurde. Zudem war sie an der Lehre der Lehrstühle von Prof. Dr. Ramin Yahyapour und Prof. Dr. Julian Kunkel beteiligt. Wir danken Frau Dr. End für ihren persönlichen Einsatz, ihre innovativen Ideen und ihre engagierte Arbeit und wünschen ihr für ihren weiteren beruflichen und privaten Lebensweg alles Gute und weiterhin viel Erfolg.

Kunkel

## NEUER MITARBEITER CHIRAG MANDAL

Seit dem 1. August 2023 ist Herr Chirag Mandal als wissenschaftlicher Mitarbeiter in der Arbeitsgruppe „Computing“ (AG C) tätig. Er machte seinen Bachelor of Technology im Jahr 2014 an der West Bengal University of Technology, Indien, in Informationstechnologie. Danach arbeitete er fünf Jahre lang in einem privaten multinationalen Unternehmen als Analyst. Ab 2019 studierte er Master of Science an der Otto-von-Guericke-Universität (OVGU) Magdeburg im Fach Data and Knowledge Engineering. In seiner Masterarbeit beschäftigte sich Herr Mandal mit der Unsicherheitsbewertung von Objekterkennung mittels Deep-Learning-Modellen. Ab 2021 arbeitete er als studentische Hilfskraft an der OVGU Magdeburg und am Fraunhofer Institut für Fabrikbetrieb und -automatisierung IFF an Anwendungen von Deep Learning auf medizinischen Bildern und LiDAR. Bei der GWGD wird er sich im Rahmen des KISSKI-Projekts mit Anwendungen des Deep Learning beschäftigen. Herr Mandal ist per E-Mail unter [chirag.mandal@gwdg.de](mailto:chirag.mandal@gwdg.de) zu erreichen.



Kunkel



## NEUER MITARBEITER MARTIN SKOWRONEK

Seit dem 1. September 2023 ist Herr Martin Skowronek Mitarbeiter in der Arbeitsgruppe „Computing“ (AG C). Als Werbekaufmann und Kulturmanager arbeitete er zunächst in Braunschweig. 2005 wechselte er nach Göttingen, wo er 2013 seinen Magister Artium in Musikwissenschaft und Philosophie an der Georg-August-Universität Göttingen machte. Bis 2020 war er als Projektmanager und Bildungszentrumsleiter beim der LEB in Niedersachsen e. V. beschäftigt und wechselte danach in die IT in den Bereichen Web Development und Cloud Migration. Herr Skowronek ist per E-Mail unter [martin.skowronek@gwdg.de](mailto:martin.skowronek@gwdg.de) zu erreichen.

Kunkel

## NEUER MITARBEITER DANIEL IMPELLIZZERI

Seit dem 1. September 2023 ist Herr Daniel Impellizzeri als studentische Hilfskraft in der Arbeitsgruppe „Anwendungs- und Informationssysteme“ (AG A) tätig. Derzeit befindet sich Herr Impellizzeri im Abschluss seines Bachelor-Studiums der Rechtswissenschaften und Volkswirtschaftslehre mit einem persönlichen Schwerpunkt auf Entwicklungsökonomik an der Georg-August-Universität Göttingen. In seinen Aufgabenbereich im Rahmen des Apple-Beratungszentrum fallen unter anderem das client- und serverseitige Device Management für macOS und iOS sowie die Unterstützung bei der Bereitstellung von Client-Backups. Für eine Vor-Ort- und Remote-Betreuung von Apple-Anwender\*innen in den Instituten und bei der GWGD ist Herr Impellizzeri per E-Mail unter [daniel.impellizzeri@gwdg.de](mailto:daniel.impellizzeri@gwdg.de) zu erreichen.



Heider

## NEUER MITARBEITER LUKAS LOWSCHIZKY

Seit dem 1. September 2023 verstärkt Herr Lukas Lowschizky das Team der Arbeitsgruppe „Anwendungs- und Informationssysteme“ (AG A) im Apple-Support der GWDG als studentische Hilfskraft. In dieser Funktion dient Herr Lowschizky als erste Anlaufstelle für Kund\*innen, die Unterstützung oder Beratung in Bezug auf Apple-Produkte benötigen. Er nimmt Anfragen entgegen, bietet direkte Hilfe an oder leitet diese zur weiteren Bearbeitung weiter. Herr Lowschizky hat sein Bachelor-Studium in Physik an der Georg-August-Universität Göttingen begonnen. Mit seinem Hintergrund, der auch einige Erfahrungen im Bereich Data Science umfasst, bringt er vielfältige Fachkenntnisse in sein Tätigkeitsfeld ein. Herr Lowschizky ist per E-Mail unter [lukas.lowschizky@gwdg.de](mailto:lukas.lowschizky@gwdg.de) zu erreichen.



Heider

## NEUE MITARBEITER\*INNEN LORENZ DAUME, INGA VON FREYTAG-LÖRINGHOFF, FELIX FUHRMANN UND ESZTER GAJDOS



Am 1. August 2023 haben Herr Lorenz Daume, Frau Inga von Freytag-Löringhoff, Herr Felix Fuhrmann und Frau Eszter Gajdos ihre Tätigkeit als studentische bzw. wissenschaftliche Hilfskraft bei der GWDG begonnen. Sie sind Teil des Support-Teams für das Office 365-Angebot für die Studierenden und Beschäftigten der Universität Göttingen. Der Office 365-Support ist sowohl per E-Mail unter [o365-support@gwdg.de](mailto:o365-support@gwdg.de) als auch in den Sprechstunden im Göttinger Rechenzentrum montags und dienstags von 9:00 – 11:00 Uhr sowie im LRC SUB am Zentralkampus mittwochs und donnerstags von 10:00 – 12:00 Uhr zu erreichen. Das Support-Team steht bei sämtlichen Fragen rund um Microsoft 365 an der Universität Göttingen zur Verfügung.

Kopp



## NEUE AUSZUBILDENDE ROEN CARKANJI UND JANNE-ILKA SCHWICKARDI

Am 1. August 2023 haben Herr Roen Carkanji und Frau Janne-Ilka Schwickardi ihre Ausbildung zum/zur Fachinformatiker\*in (IHK) in der Fachrichtung Systemintegration bei der GWDG begonnen. Beide gehören der Arbeitsgruppe „Nutzerservice und Betriebsdienste“ (AG H) an.

Herr Carkanji besuchte bis zu seinem Ausbildungsbeginn das berufliche Gymnasium in der Fachrichtung IT an der BBS II in Göttingen und sammelte dort erste Erfahrungen im Bereich Informatik. Aufgrund des stärkeren Praxisbezugs entschied er sich nun für eine Ausbildung bei der GWDG. Herr Carkanji ist per E-Mail unter [roen.carkanji@gwdg.de](mailto:roen.carkanji@gwdg.de) zu erreichen.

Frau Schwickardi arbeitet bereits seit September 2018 bei der GWDG und war bis zum Beginn ihrer Ausbildung als studentische Hilfskraft im 1st-Level-Support der Service-Hotline der GWDG bzw. seit Februar dieses Jahres als Aushilfskraft in der Arbeitsgruppe „IT-Infrastruktur“ (AG I) tätig. Aufgrund ihrer umfangreichen Vorbildung startet sie ihre Ausbildung im zweiten Lehrjahr. Frau Schwickardi ist per E-Mail unter [janne-ilka.schwickardi@gwdg.de](mailto:janne-ilka.schwickardi@gwdg.de) zu erreichen.



Kopp



# Software und Lizenzverwaltung

## Der einfache Weg zur Software!

### Ihre Anforderung

Sie benötigen eine Software, für die es keine von Ihnen nutzbare Rahmenvereinbarung gibt. Die Anzahl der erforderlichen Lizenzen ist nicht genau festgelegt.

### Unser Angebot

Wir verfügen über eine Reihe von Rahmen- und Campusvereinbarungen mit namhaften Softwareherstellern und -lieferanten, über die Software auch in geringerer Stückzahl bezogen werden kann. Wir wickeln für Sie die Beschaffung der erforderlichen Lizenzen ab. Wir können uns bei Vertragsverhandlungen und Bedarfsanalysen engagieren. Zugriffslizenzen können auch über Lizenzserver verwaltet werden.

### Ihre Vorteile

> Sie können die benötigte Software in vielen Fällen sofort nutzen.

- > Sie brauchen kein eigenes Ausschreibungs- und Beschaffungsverfahren durchzuführen.
- > Sie ersparen sich die zeitraubenden Verhandlungen mit den Softwareherstellern und -lieferanten.
- > Die Anzahl der benötigten Lizenzen wird Ihnen flexibel zur Verfügung gestellt.
- > Wir können die Nachfrage von verschiedenen Nutzer\*innen für neue Lizenzvereinbarungen bündeln.

### Interessiert?

Informationen zu bestehenden Lizenzvereinbarungen sind auf der u. g. GWDG-Webseite zu finden. Falls Sie nach spezieller Software suchen, die noch nicht auf unserer Webseite erwähnt ist, kommen Sie bitte auf uns zu. Wir werden prüfen, ob wir eine Vereinbarung abschließen können und bündeln die Nachfrage mit anderen Nutzer\*innen.

>> [www.gwdg.de/software](http://www.gwdg.de/software)



**INFORMATIONEN:**  
support@gwdg.de  
0551 39-30000

November bis  
Dezember 2023

# Academy

KURS	DOZENT*IN	TERMIN	ANMELDEN BIS	AE
DEEP LEARNING WITH GPU CORES	Sommer, Meisel, Kirchner	07.11.2023 9:30 – 13:00 Uhr	31.10.2023	2
GRUNDLAGEN DER BILDBEARBEITUNG MIT AFFINITY PHOTO	Töpfer	08.11. – 09.11.2023 9:30 – 16:00 Uhr	01.11.2023	8
WORKING WITH GRO.DATA	Dr. Király	14.11.2023 10:00 – 11:30 Uhr	13.11.2023	0
EINFÜHRUNG IN DIE STATISTISCHE DATENANALYSE MIT SPSS	Cordes	15.11. – 16.11.2023 9:00 – 12:00 und 13:00 – 15:30 Uhr	08.11.2023	8
ADMINISTRATION IM ACTIVE DIRECTORY	Quentin, Hast, Kopp	22.11.2023 9:00 – 12:30 und 13:30 – 15:30 Uhr	15.11.2023	4
MONITORING WITH TIG STACK IN THE GWDC	Merz	23.11.2023 9:00 – 10:30 Uhr	16.11.2023	1
HYBRID IDENTITY – INTEGRATION DES ACTIVE DIRECTORY IN MICROSOFT AZURE ACTIVE DIRECTORY	Kopp	28.11. – 29.11.2023 9:00 – 12:00 und 13:00 – 16:00 Uhr	21.11.2023	8
RECHENZENTRUMSTECHNIK – KURS FÜR AUSZUBILDENDE	Schumann	29.11. – 30.11.2023 9:00 – 16:00 Uhr	22.11.2023	8
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VORWISSEN	Cordes	06.12. – 07.12.2023 9:00 – 12:00 und 13:00 – 15:30 Uhr	29.11.2023	8
VIRTUELLE CLOUD-INFRASTRUKTUREN – KURS FÜR AUSZUBILDENDE	Kopp	12.12. – 14.12.2023 9:00 – 16:00 Uhr	05.12.2023	12
WORKING WITH GRO.DATA	Dr. Király	12.12..2023 10:00 – 11:30 Uhr	11.12.2023	0

#### Teilnehmerkreis

Das Angebot der GWGD 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 GWGD 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 im Kundenportal der GWGD (<https://www.gwdg.de>) einloggen. Wenn Sie zum Kreis der berechtigten Nutzer\*innen der GWGD gehören und noch keinen GWGD-Account besitzen, können Sie sich im Kundenportal unter dem URL <https://www.gwdg.de/registration> registrieren. 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 GWGD – in das interne Kosten- und Leistungsrechnungssystem der GWGD 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 GWGD 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.



# GWGD 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 Schreibberechtigung 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 CodiMD 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 „GWGD 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 GWGD-Accounts oder die einmalige Registrierung unter <https://www.gwdg.de/registration> oder erforderlich.

>> [www.gwdg.de/gwdgpad](https://www.gwdg.de/gwdgpad)



Gesellschaft für wissenschaftliche  
Datenverarbeitung mbH Göttingen