GWDG NACHRICHTEN 04|21

Big Data and HPC

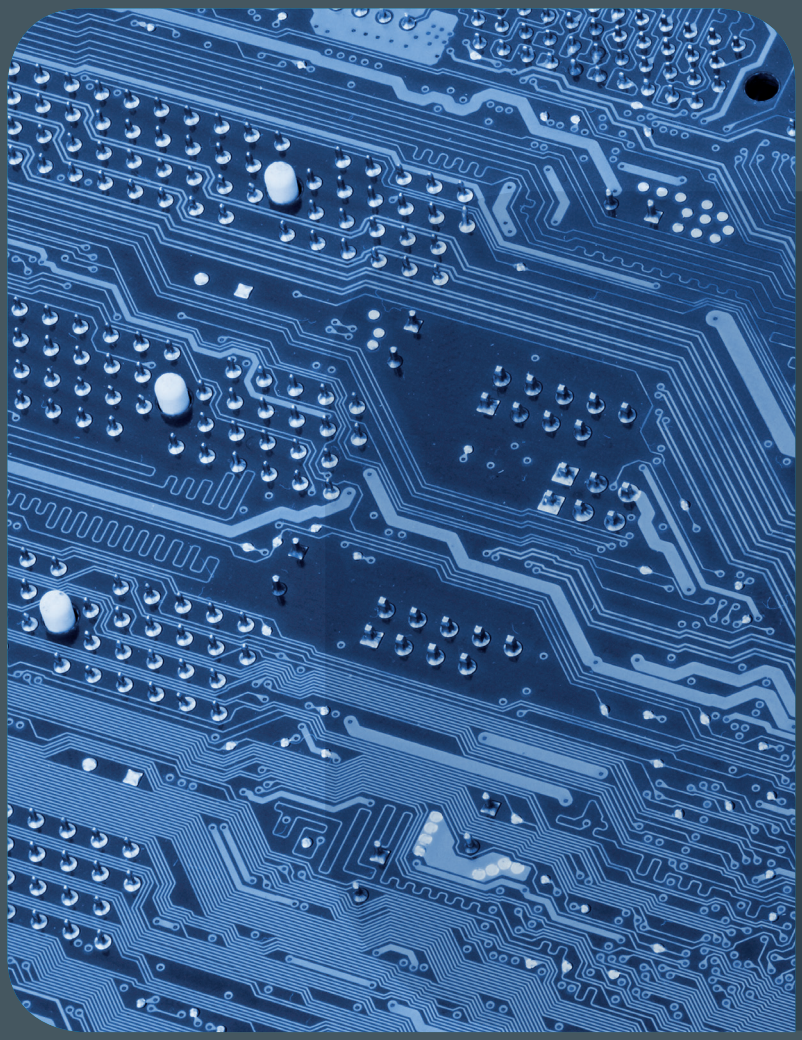
Nutzerzertifikate in der DFN-PKI

Kubernetes with Rancher

Persistent Identifier

ZEITSCHRIFT FÜR DIE KUNDEN DER GWDG

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen



GWDG
NACHRICHTEN

04|21

Inhalt

Impressum

Zeitschrift für die Kunden der GWDG

ISSN 0940-4686
44. Jahrgang
Ausgabe 4/2021

Erscheinungsweise:
10 Ausgaben pro Jahr

www.gwdg.de/gwdg-nr

Auflage:
550

Fotos:
© profit_image - stock.adobe.com (1)
© pineapple - Fotolia.com (19)
© contrastwerkstatt - Fotolia.com (24)
© Robert Kneschke - Fotolia.com (26)
© MPIbpc-Medienservice (3, 25)
© GWDG (2, 23)

Herausgeber:
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
Am Faßberg 11
37077 Göttingen
Tel.: 0551 201-1510
Fax: 0551 201-2150

Redaktion:
Dr. Thomas Otto
E-Mail: thomas.otto@gwdg.de

Herstellung:
Maria Geraci
E-Mail: maria.geraci@gwdg.de

Druck:
Kreationszeit GmbH, Rosdorf

Prof. Dr. Ramin Yahyapour
ramin.yahyapour@gwdg.de
0551 201-1545

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

es ist bekannt, dass viele URL-Referenzen bereits nach einigen Jahren nicht mehr zu der gewünschten Information führen. Für wissenschaftliche Veröffentlichungen ist dies ein gravierendes Problem, welches die Nachnutzung und Reproduzierbarkeit von Forschung erschwert.

Persistenten Identifikatoren, sogenannten PIDs, kommt daher eine große Bedeutung zu, um die korrekte Auflösung von solchen Referenzen langfristig zu gewährleisten. Sollten sich der Speicherort oder Metadaten von Informationen ändern, so können diese im PID aktualisiert werden. Wenn die referenzierten Daten zudem in einem geeigneten Repositorium gespeichert sind, können Änderungen auch automatisch gepflegt werden. Persistente Identifikatoren sind daher ein wichtiger Bestandteil für jedes Forschungsdatenmanagement.

Die GWDG bietet hier bereits seit einigen Jahren einen PID-Dienst an und ist im Rahmen des internationalen ePIC und DONA-Konsortiums ein Hosting-Knoten für PIDs in Deutschland. In dieser Ausgabe der GWDG-Nachrichten berichten wir über dieses Angebot und wie es genutzt werden kann. Die GWDG arbeitet in Kooperation mit weiteren Partnern daran, diese Dienste ständig weiterzuentwickeln. Daher werden wir auch weiterhin immer wieder an dieser Stelle darüber berichten.

Ramin Yahyapour

GWDG – IT in der Wissenschaft

Big Data and High Performance Computing

Text and Contact:
Dr. Jack Ogaja
jack.ogaja@gwdg.de
0551 39-30118

As a provider of powerful and innovative IT services, the GWDG provides its users with extensive High Performance Computing (HPC) resources to solve a wide range of problems in the field of Data Science and related topics. This article gives a brief overview of how HPC systems drive Big Data Analytics. It also briefly describes the resources available in the GWDG Scientific Compute Cluster (SCC) and a planned GWDG Academy course on High Performance Data Analytics.

WHAT IS BIG DATA?

Big Data is often defined by its extreme volume, the many varieties of data types it contains and the fast rate or velocity at which it must be processed. Considering the advancement in computer and mobile technologies and their inevitable effects on human social behaviour, Big Data has become integral part of human activities from social networking to online shopping, and is expected to grow tremendously in the coming years (see figure 1). Big Data Analytics is thus ubiquitous in scientific research, industrial production and business services. To reap the benefits of Big Data, Big Data Analytics demands advance and powerful computer processing technologies which can handle its unique characteristics.

BIG DATA HPC CONVERGENCE

HPC traditionally harnesses the power of parallel compute nodes and distributed storage system to speed up computer workloads. The compute nodes often include powerful CPUs, GPUs or a hybrid of both with fast memories making intensive data processing more efficient. At the centre of Big Data HPC convergence, scalable novel tools, frameworks and platforms are used to gain new insights, and discover patterns and trends from massive data sets through parallel and distributed processes. This is often referred to as *High Performance Data Analytics (HPDA)*. *Machine Learning (ML)* and *Deep Learning (DL)* which use sophisticated algorithms to extract useful information from massive data sets form part of HPDA.

THE COMPLEX DATA ARCHITECTURE FOR DATA ANALYTICS

Central to Data Analytics is data architecture. A typical data architecture consist of three main areas: data sources, where data is generated and collected; data storage, where generated and collected data are stored and processed; and applications, where data are shared with consumers. In the data sources and storage areas, distinctions can be drawn between Small (traditional) and Big Data architectures which result to different storage and programming models for Data Analytics workflows.

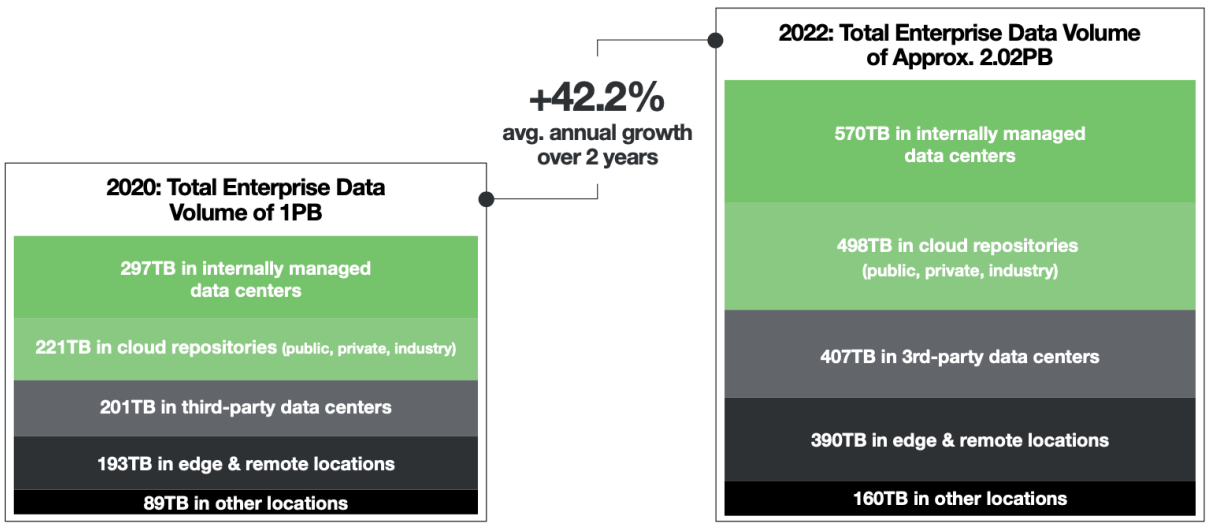
For Small Data, *Relational Database Management Systems (RDMS)* are commonly used for data integration and storage. Over the RDMS software, *Online Transaction Processing (OLTP)* can be used to generate and process data. HPC resources can be used to integrate transactional or operational Small Data from multiple systems to support operational reporting in what is known as an *Operational Data Store (ODS)* system. The same applies to *Massively Parallel Processing (MPP)* database in which data is partitioned across multiple HPC servers and each server can process the data on that server locally and independently. A *Data Warehouse (DW)* i.e. a well-designed and centralized data depository, can play an important role in reducing the time it takes to obtain results from a data analytics workflow. In a DW, data from multiple databases can be merged. Summaries of historic data can then be generated through *Online Analytical Processing (OLAP)*.

For Big Data, *Hadoop* and *Spark* are commonly used open-source frameworks for processing massive amounts of data. They are both developed by the Apache Software Foundation. Hadoop includes a scalable distributed file system where data files can be stored across multiple servers, i.e. servers, storage and compute components can be added to accommodate an increasing volume of data. To process data, Hadoop uses *MapReduce* – a programming model and technique which uses parallel and distributed

Big Data und Hochleistungsrechnen

Als Anbieter leistungsfähiger und innovativer IT-Dienstleistungen stellt die GWDG ihren Nutzer*innen umfangreiche High Performance Computing (HPC)-Ressourcen zur Lösung von vielfältigen Fragestellungen im Bereich Data Science und damit verwandten Themen zur Verfügung. Dieser Artikel gibt einen kurzen Überblick darüber, wie HPC-Systeme Big Data Analytics vorantreiben. Darüber hinaus werden die im Scientific Compute Cluster (SCC) der GWDG verfügbaren Ressourcen sowie ein geplanter Kurs der GWDG Academy zu High Performance Data Analytics kurz beschrieben.

FIGURE 1 **Expected Annual Data Growth Rate**



Source: The Seagate Rethink Data Survey[1], IDC, 2020

Figure 1: Expected annual data growth rate (Footnote 1: The Seagate Rethink Data Survey by IDC is IDC's name for the survey whose findings are discussed in the Seagate Technology report "Rethink Data: Put More of Your Business Data to Work – From Edge to Cloud")

algorithm on a HPC or distributed system. Spark which does not provide a distributed file system instead uses *Resilient Distributed Datasets (RDD)* as the primary API. With RDD, immutable distributed collection of data elements partitioned across HPC nodes allows parallel operations. Spark can thus be used together with Hadoop to create distributed datasets from files stored in Hadoop's distributed file system. Proprietary Data Analytics tools also exist e.g. Intel's *oneAPI Toolkit* for *Artificial Intelligence (AI)* which also includes Intel Distribution for Python.

In the application area, there are many ML and DL libraries, frameworks and tools both open-source and proprietary. Some commonly used frameworks and libraries are: *Tensorflow* which is an ML library with a particular focus on training and inference of deep neural networks and is primarily developed by Google Brain Team, *PyTorch* is an open-source ML library primarily developed by Facebook's AI Research Lab, and *Anaconda*, a distribution of Python and R programming languages for scientific computing by Anaconda Inc.

HPDA RESOURCES IN THE GWDG SCIENTIFIC COMPUTE CLUSTER

HPDA resources (hardware and software components) are readily available for users in the GWDG Scientific Compute Cluster (SCC). The hardware component includes CPU and GPU nodes, and a well integrated storage system. A total of 138 GPU cards distributed in 39 nodes are available (see [1]). The GPU nodes consist of: 2 nodes with 8 x Tesla V100 each; 14 nodes with 4 x Quadro RTX5000 each; 7 nodes with 2 x GTX 1080 each; 8 nodes with 2 x GTX980 each; and 10 nodes with 2 x Tesla K40m each. The software component includes some of the most powerful open-source tools, libraries and frameworks mentioned above e.g. *Apache Spark*, *Tensorflow*, and *Jupyterhub*. Additional information can be found on the GWDG-SCC website i.e. see links [2], [3], and [4]. The resources are available for users with GWDG accounts.

INTRODUCTORY COURSE ON HPDA AT THE GWDG ACADEMY

Developing and maintaining efficient tools for processing and analysing Big Data in HPC systems is necessary for discovering patterns and gaining insights for data-intensive topics including bimolecular science, global climate change, accurate weather prediction, cancer research and cybersecurity among others. Building enough man-power (human resources) to be able to utilize the increasing computational power in HPC infrastructure to process and analyse Big Data is of great importance in advancing Big Data Analytics and ML.

For this reason, at the GWDG Academy, a new introductory course on HPDA is planned for the second-half of 2021 with an objective of providing interested learners with foundational knowledge on emerging tools for Data Analysis in HPC systems. Participants will learn how to organize data in Data Lakes and DWs, and create massive data pipelines to obtain meaningful information from their data. This course is targeted at researchers and students using the HPC system for data-intensive experiments.

Different topics covered in the course include introduction to HPC for Data Analytics, introduction to Data Architectures and massive data pipelines, novel softwares and tools for HPC Data Analytics, and introduction to Machine and Deep Learning.

LINKS

[1] *https://www.gwdg.de/hpc-on-campus/scc*
[2] *https://info.gwdg.de/docs/doku.php?id=en:services:application_services:high_performance_computing:spark*
[3] *https://info.gwdg.de/docs/doku.php?id=en:services:application_services:high_performance_computing:tensorflow*
[4] *https://info.gwdg.de/docs/doku.php?id=en:services:application_services:jupyter:start*

Weitere Überarbeitung des Beantragungsweges für Nutzerzertifikate in der DFN-PKI

Text und Kontakt:
Thorsten Hindermann
thorsten.hindermann@gwdg.de
0551 201-1837

Nachdem der DFN-Verein den im Herbst 2019 eingeführten neuen Beantragungsweg für Nutzerzertifikate im Frühjahr 2020 überarbeitet hat, beschrieben in den GWDG-Nachrichten 4-5/2020, ist nun seit Mitte Februar dieses Jahres eine nochmals überarbeitete Beantragungsseite für Nutzerzertifikate vom DFN-Verein freigegeben worden. Diese Seite wird in diesem Artikel vorgestellt.

BEGRIFFSERKLÄRUNGEN

Die zwei Hauptbegiffe, die im Zusammenhang mit dem Umgang von E-Mail-Verschlüsselung fallen, sind X.509-Zertifikate und Public Key Infrastructure, im Weiteren kurz PKI genannt. Die PKI ist ein hierarchisch organisierter Aufbau von Zertifikatsautoritäten, engl. Certification Authorities (im Weiteren kurz CAs genannt), beginnend mit einer Wurzel über Zwischenstationen hin zur ausstellenden Autorität für Zertifikate. Diese Kette der Autoritäten bildet die Grundlage einer PKI.

Die Zertifikate wiederum sind eine digitale Repräsentation von Benutzer*innen, Diensten, Netzwerkgeräten oder Computern, die durch eine CA ausgestellt wurden. Diese Zertifikate sind zusammen mit jeweils einem privaten Schlüssel (engl. private key) und einem öffentlichen Schlüssel (engl. public key) miteinander verbunden.

Technisch betrachtet ist das Zertifikat eine digital signierte Ansammlung von Informationen, u. a. Informationen über den /die Benutzer*in, den Dienst, das Netzwerkgerät oder den Computer, die ausstellende CA, die verwendeten Signier-/Verschlüsselungsverfahren, die Abruf-URLs von Sperrlisten für gesperrte Zertifikate usw. X.509 wiederum ist ein ITU-T-Standard (Internationale Fernmeldeunion) für eine PKI zum Er-/Ausstellen digitaler Zertifikate.

ZERTIFIKAT BEANTRAGEN

Um nun ein Zertifikat zu beantragen, ist es als erstes wichtig zu wissen, welche ausstellende Registrierungsautorität, engl. Registration Authority (im Weiteren kurz RA genannt), für Antragsteller zuständig ist. Unter [1] finden Sie die jeweiligen Einstiegspunkte für die RAs der Max-Planck-Gesellschaft, der Universität Göttingen und der GWDG.

In den GWDG-Nachrichten 4-5/2020 wurde der überarbeitete Beantragungsweg erstmalig ausführlich dargestellt. Durch die im Februar 2021 durch den DFN-Verein vorgenommene nochmalige Überarbeitung ändert sich der Abschnitt „Zertifikat beantragen“, den Sie unter [2] finden.

Die Einstiegsseite sieht wie in Abbildung 1 dargestellt aus.

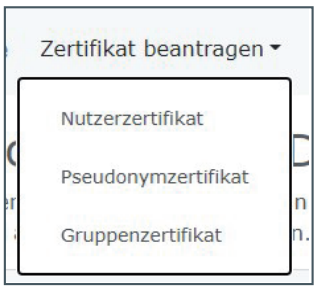


Abb. 2

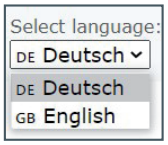


Abb. 3

Abb. 5

Auffällig sind die drei großen Kacheln für ein Nutzer-, Pseudonym- und Gruppenzertifikat. Darunter zwei kleinere Kacheln, eine für das Abholen des ausgestellten Zertifikats und daneben die Möglichkeit der Sperrung eines Zertifikats. Darüber hinaus gibt es eine Menüzeile mit den Menüpunkten „Zertifikat beantragen“, das wiederum ein Ausklapp-Menü ist und die Beantragungsmöglichkeiten für die drei Zertifikattypen enthält, die mit den drei großen Kacheln korrespondieren (siehe Abbildung 2). Daneben gibt es die beiden Menüpunkte „Zertifikat abholen“ und „Zertifikat sperren“, die wiederum mit den beiden kleineren Kacheln korrespondieren. Rechts außen gibt es einen Sprachumschalter von Deutsch auf Englisch (siehe Abbildung 3). Das Ergebnis dieser Aktion, dem Umschalten der Sprache, ist in Abbildung 4 zu

Further Revision of the Application Procedure for User Certificates in the DFN-PKI

After the DFN-Verein has revised the new application procedure for user certificates introduced in fall 2019 in spring 2020, described in the GWDG News 4-5/2020, a further revised application page for user certificates has now been released by the DFN-Verein since mid-February this year. This page is presented in this article.

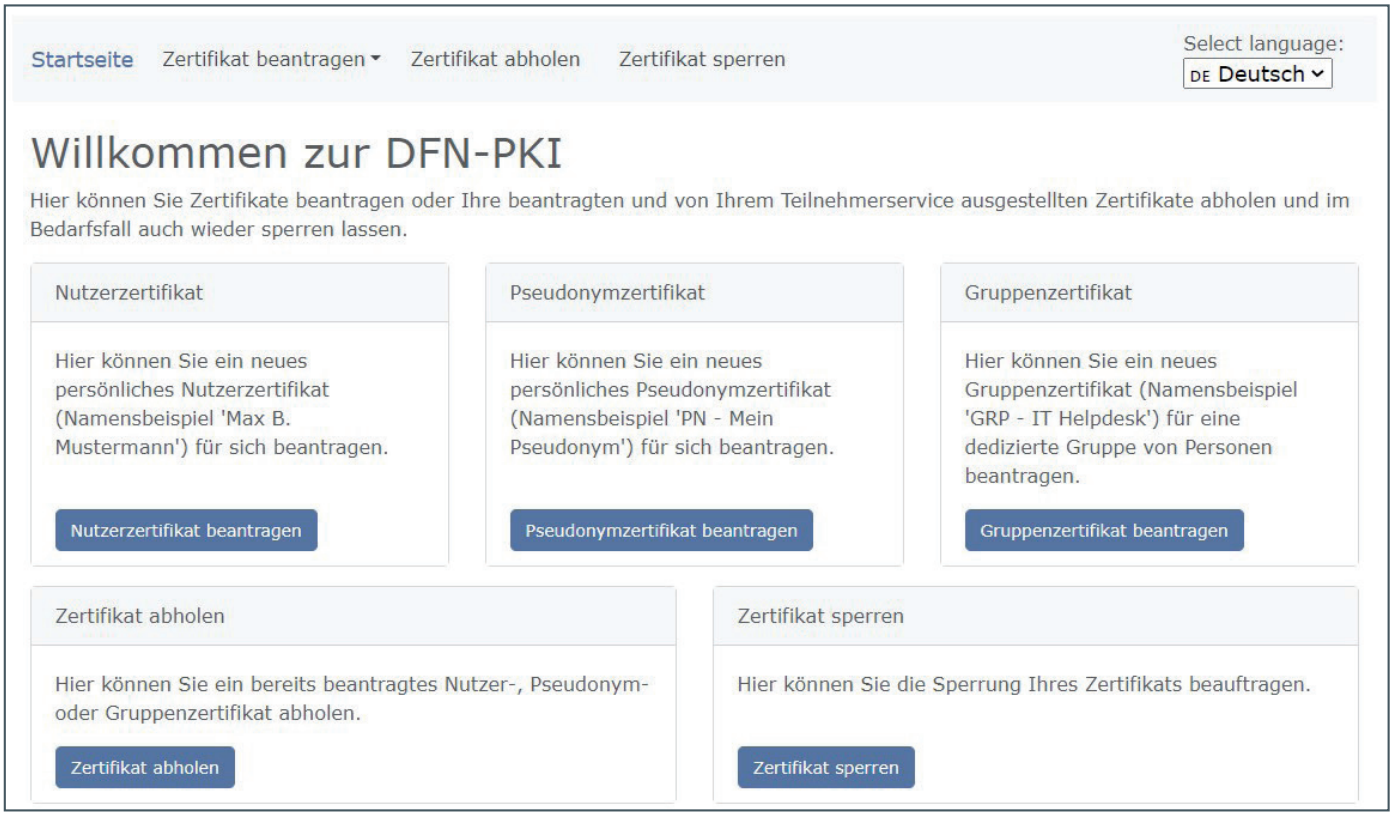


Abb. 1

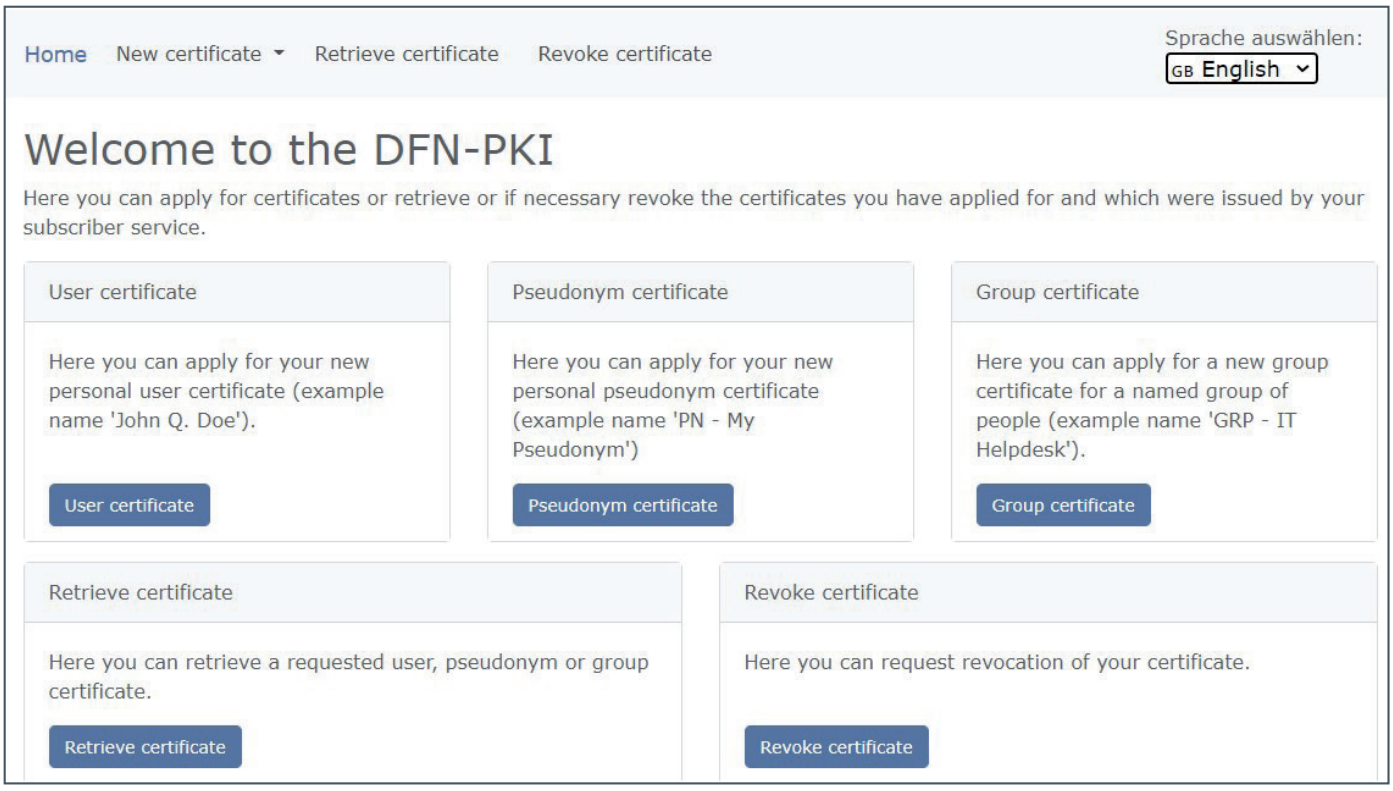


Abb. 4

sehen. In den folgenden Abschnitten werden diese fünf „Kacheln" bzw. Menüpunkte näher beschrieben.

Ein wichtiger Hinweis vorab: Die Möglichkeit, mit dem Webbrowser „Microsoft Internet Explorer" Nutzerzertifikate in der DFN-PKI zu beantragen (siehe auch die GWDG-Nachrichten 12/2019, Seite 10, [3]), steht seit dem 01.03.2021 nicht mehr zur Verfügung. Sie können daher für die Zertifikatbeantragung nur noch die bekannten Webbrowser wie z. B. Microsoft Edge, Mozilla Firefox oder Google Chrome nutzen. Ausführliche Informationen zur Beantragung von Nutzerzertifikaten finden Sie unter [1].

Nutzerzertifikat

Mit einem Klick auf die Schaltfläche „Nutzerzertifikat beantragen" wird die Beantragung eines persönlichen Nutzerzertifikats gestartet (siehe Abbildung 5).

In dem neuen Formular ist das vorherige „Namen

Abb. 6

(CN)"-Eingabefeld nun aufgeteilt in „Vorname (GN)" und „Nachname (SN)". Vorangestellt ist ein Eingabefeld für Namenszusätze, so wie diese in amtlichen Ausweisdokumenten eingetragen sind (z. B. „Dr."). Zusätzlich sind nur in dem Antragsformular für die GWDG noch zwei Eingabefelder für alternative E-Mail-Adressen vom DFN-Verein konfiguriert worden (siehe Abbildung 6).

Die übrigen Eingabefelder sind zur vorherigen Version unverändert geblieben. Mit einem Klick auf die Schaltfläche „Weiter" (siehe Abbildung 7) geht es weiter zur Zusammenfassungsseite.

Auf der Zusammenfassungsseite (siehe Abbildung 8) können die eingegebenen Daten überprüft werden. Mit einem Klick auf die Schaltfläche „Daten ändern" können die Daten noch geändert werden. Mit einem Klick auf „Antragsdatei speichern" wird ein Dialog zur Eingabe eines Passwortes für die zu speichernde Antragsdatei angezeigt (siehe Abbildung 9). Mit einem Klick auf die Schaltfläche „OK" wird die Antragsdatei im Standard Downloadordner des verwendeten Webbrowsers gespeichert. In diesem Beispiel unter *P:\Downloads\Antragsdatei_Thorsten_Hindermann_78243104_2021-03-11.json*. Der Dateiname variiert je nach Antragstyp und Angaben im Zertifikat. Diese Datei bitte nicht löschen und gut aufbewahren!

Die letzte Seite zeigt die Möglichkeit, die Antragsdatei noch einmal über die Schaltfläche „Antragsdatei (JSON) erneut speichern" zu speichern. Am wichtigsten ist nun aber, den Antrag mit Klick auf die Schaltfläche „Zertifikatantragsformuler (PDF) herunterladen" herunterzuladen (siehe Abbildung 10). Die PDF-Datei wird im Standard Downloadordner des verwendeten Webbrowsers gespeichert. In diesem Beispiel unter *P:\Downloads\ Zertifikatantrag_Thorsten_Hindermann_78243104_2021-03-11.pdf*. Der Dateiname variiert je nach Antragstyp und Angaben im Zertifikat. Das weitere Vorgehen wird im späteren Abschnitt „Weitere Schritte" beschrieben.

Pseudonymzertifikat

Mit einem Klick auf die Schaltfläche „Pseudonymzertifikat beantragen" wird die Beantragung eines persönlichen Pseudonymzertifikats gestartet (siehe Abbildung 11).

Ein Pseudonymzertifikat ist immer dann zu wählen und zu beantragen, wenn im Zertifikat Angaben gemacht werden, die nicht in einem amtlichen Ausweisdokument stehen. Im Beispiel von Abbildung 12 ist durch den Zusatz in Klammern die Notwendigkeit eines Pseudonymzertifikatantrags gegeben.

Die in Abbildung 13 angezeigten Daten sind Kontaktdaten, die aber nicht mit im Zertifikat selber aufgenommen werden. Die abschließenden Schritte nach dem Klick auf die Schaltfläche „Weiter" sind in den beiden letzten Absätzen des vorherigen Abschnitts „Nutzerzertifikat" beschrieben. Das weitere Vorgehen wird im späteren Abschnitt „Weitere Schritte" beschrieben

Gruppenzertifikat

Mit einem Klick auf die Schaltfläche „Gruppenzertifikat beantragen" wird die Beantragung eines Gruppenzertifikats für eine

Abb. 7

Gruppe von Personen gestartet (siehe Abbildung 14).

Ein Gruppenzertifikat ist immer dann zu wählen, wenn mehrere Personen über eine gleiche E-Mail-Adresse E-Mails versenden, z. B. ein IT-Servicedesk, ein Sekretariat etc. Im Beispiel von Abbildung 15 erfolgt die Beantragung für eine CA oder auch RA, da es im Idealfall mindestens zwei oder mehr Personen gibt, eine ein CA oder RA verwalten.

Die in Abbildung 16 angezeigten Daten sind Kontaktdaten, die aber nicht mit im Zertifikat selber aufgenommen werden. Die abschließenden Schritte nach dem Klick auf die Schaltfläche „Weiter" sind in den beiden letzten Absätzen des früheren Abschnitts „Nutzerzertifikat" beschrieben.

WEITERE SCHRITTE

Die PDF-Datei, die in einem der vorherigen Schritte heruntergeladen worden ist, ist nun mit einem PDF-Programm zu öffnen, auszudrucken und zu unterschreiben. Vor der COVID-19-Pandemie konnte der Antrag persönlich beim Teilnehmerservice im eigenen Institut abgegeben werden. Dies kann in der COVID-19-Zeit sicherlich auch erfolgen, wenn das Institut schon wieder teilweise oder komplett in Präsenzarbeit vertreten ist – in diesem Fall dann aber unter Einhaltung der Hygienemaßnahmen des betreffenden Instituts. Im anderen Fall, wenn die meisten Anwender*innen im Homeoffice arbeiten, muss dem Teilnehmerservice die unterschriebene PDF-Datei oder das Foto des unterschriebenen Antrags z. B. als JPEG-Datei per E-Mail zugesendet werden. Der TS-MA druckt den Antrag aus und vereinbart einen Termin für eine persönliche Identifizierung per Videokonferenz. Dieses Verfahren ist in den GWDG-Nachrichten 12/2020 auf Seite 11 [4] im Abschnitt „Persönliche Identifizierung" ausführlich beschrieben. Nach erfolgreicher persönlicher Identifizierung und vergleichender Prüfung wird der zuständige TS-MA den Zertifikatantrag genehmigen. Nach der Genehmigung eines der drei oben beschriebenen Zertifikattypen wird eine Bestätigungs-E-Mail an die beantragende Person gesendet.

Für die Abholung und Fertigstellung eines der oben beschriebenen Zertifikattypen beachten sie bitte den folgenden Abschnitt.

ZERTIFIKAT ABHOLEN

Mit einem Klick auf die Schaltfläche „Zertifikat abholen" können beantragte und ausgestellte Nutzer-, Pseudonym- oder Gruppenzertifikate abgeholt und fertiggestellt werden (siehe Abbildung 17). Die Abholung des ausgestellten Zertifikats wird in den GWDG-Nachrichten 4-5/2020 auf Seite 28 [5] im Abschnitt „Abholung" beschrieben.

ZERTIFIKAT SPERREN

Mit einem Klick auf die Schaltfläche „Zertifikat sperren" wird die Sperrung eines gültigen Zertifikats gestartet (siehe

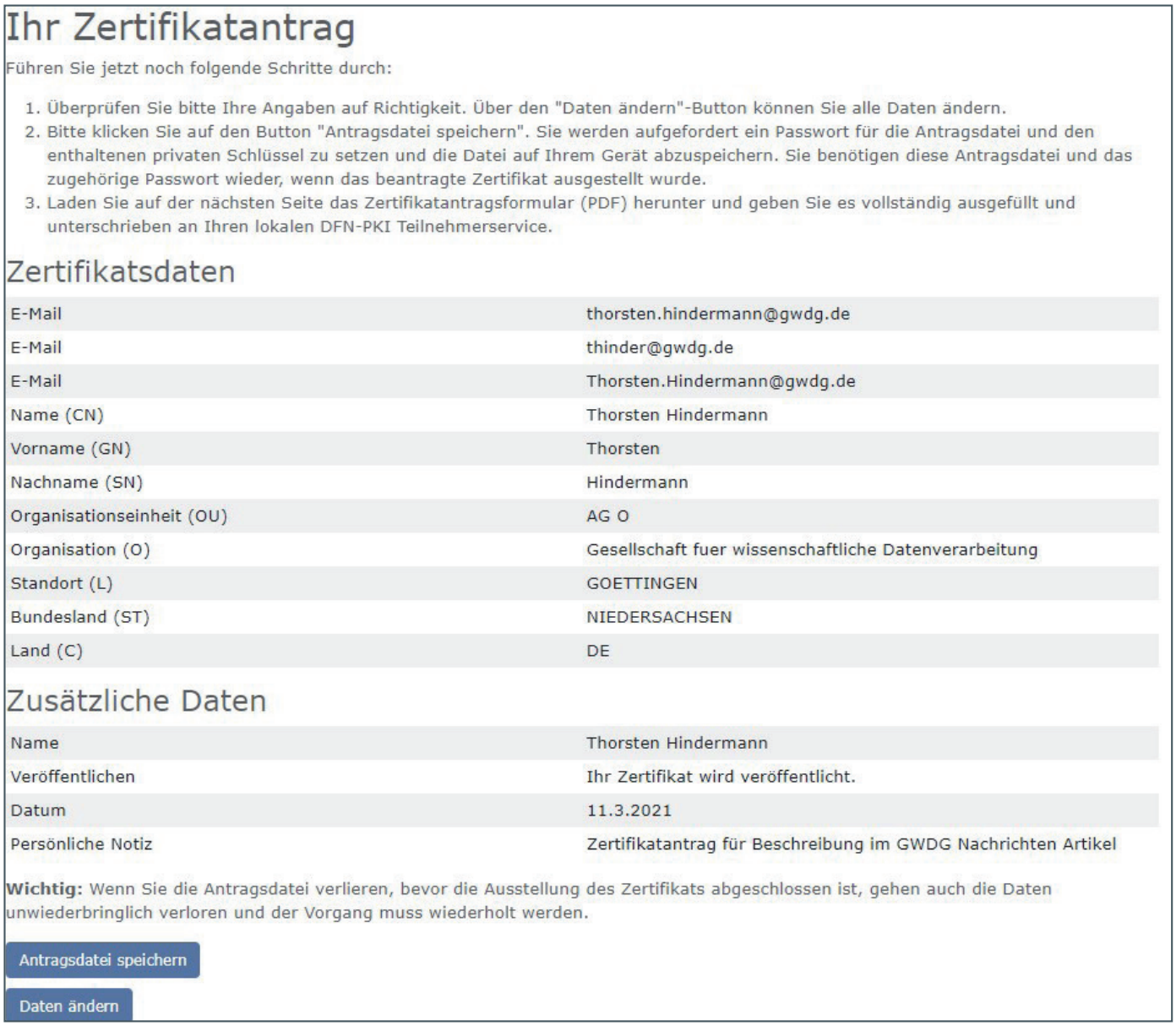


Abb. 8

Abb. 9

Abbildung 18).

Sofern Sie bei der Ausstellung des zu sperrenden Zertifikats eine E-Mail mit dem Zertifikat bekommen haben, befindet sich die Seriennummer in dieser E-Mail. Ansonsten kann die Seriennummer des zu sperrenden Zertifikats und dessen Zertifikatdetails in der Zertifikatanzeige Ihres jeweils genutzten Betriebssystems bzw.

Abb. 11

Abb. 14

Ihrer Anwendungssoftware, wie z. B. Ihr E-Mail-Programm, ausgegeben werden.

Der besagte Ausschnitt aus der E-Mail sieht wie in Abbildung 19 dargestellt aus. Zum schnelleren Auffinden ist in diesem Beispiel die Seriennummer fett markiert (Anmerkung des Autors: Die Beispiel-Zertifikate in diesem Artikel sind inzwischen alle gesperrt.).

Die Sperr-PIN wurde bei der Beantragung des Zertifikats von dem/der Beantragenden selbst gesetzt. Falls Sie die Sperr-PIN nicht mehr wissen, wenden Sie sich bitte an Ihren lokalen Teilnehmerservice in Ihrem Institut.

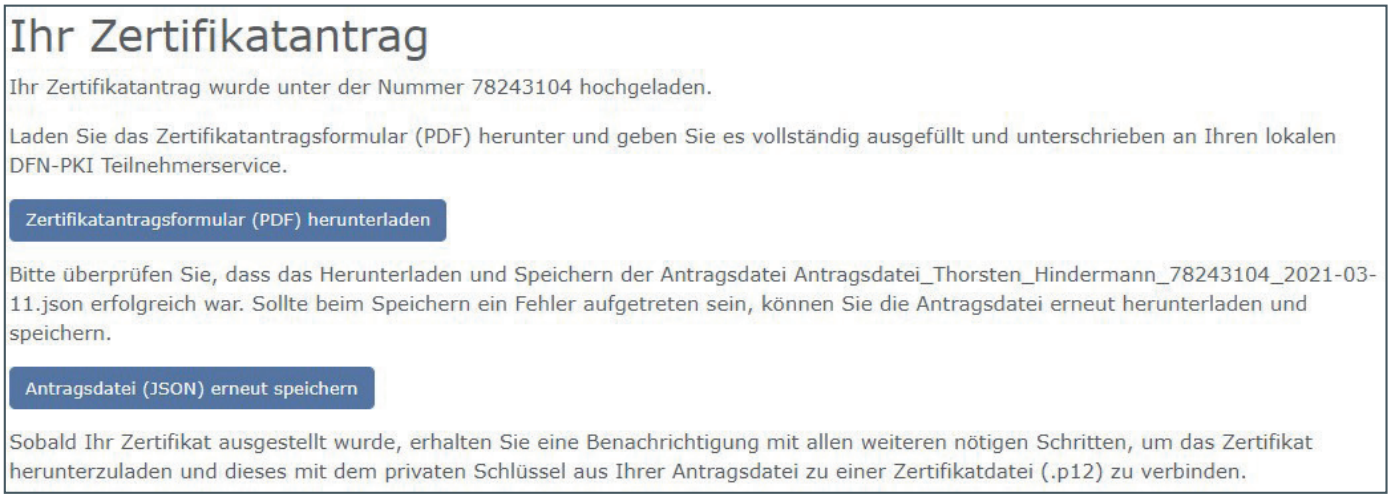


Abb. 10

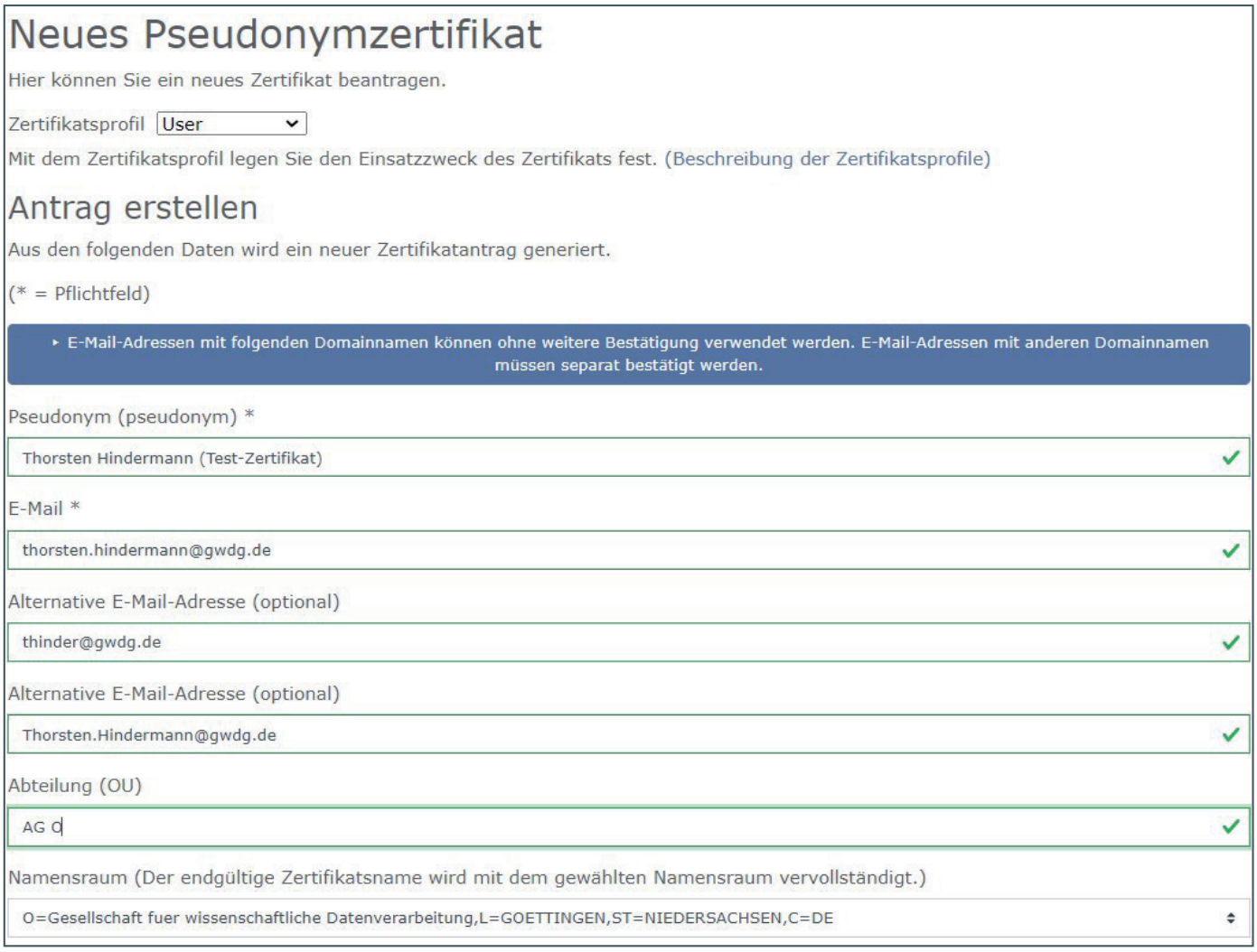


Abb. 12

Beim Sperrgrund kann aus seiner Liste von Gründen gewählt werden (siehe Abbildung 20). Zu den ausgewählten Sperrgründen werden auf dem Formular folgende ergänzende Erklärungen angezeigt (siehe Abbildung 21). Der fertig ausgefüllte Zertifikatsperrantrag sieht dann wie in Abbildung 22 dargestellt aus.

Mit einem Klick auf die Schaltfläche „Zertifikat sperren“ ist der Sperrantrag eingereicht und wird vom zuständigen Teilnehmerservice des Instituts bearbeitet. Weitere Detailinformationen zur Zertifikatsperrung können Sie den GWDG-Nachrichten 12/2020 auf Seite 17 [6] im Abschnitt „Sperrung“ und den GWDG-Nachrichten 3/2021 auf Seite 11 [7] im Abschnitt „Sperranträge“ entnehmen.

WEITERE INFORMATIONEN

In diesem Artikel haben die Antragsteller*innen eines der drei beschriebenen Zertifikattypen Informationen und Anleitungen erhalten, um erfolgreich ein Zertifikat zu beantragen. Die weiteren Schritte mit dem ausgestellten und fertigen Zertifikat können in

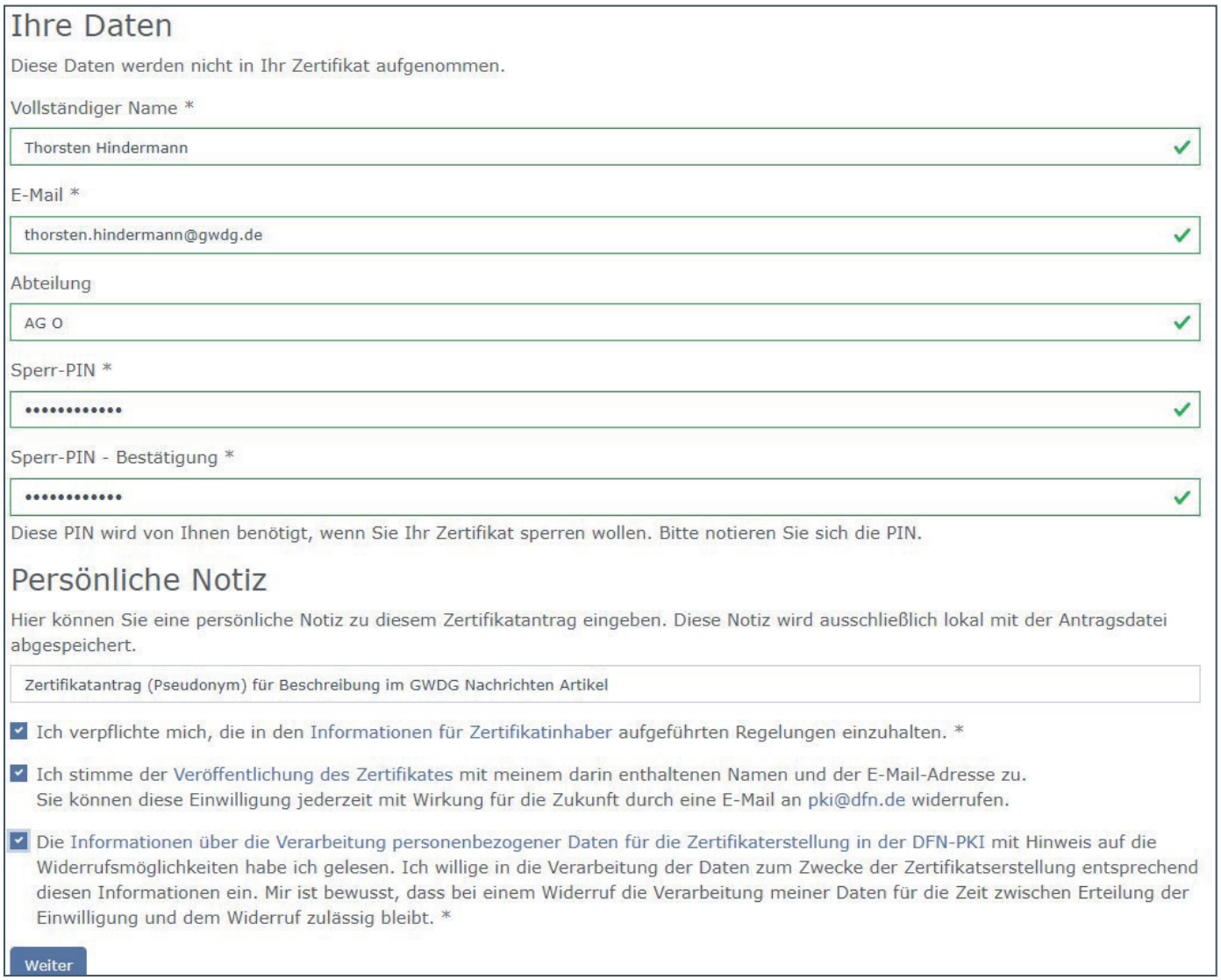


Abb. 13

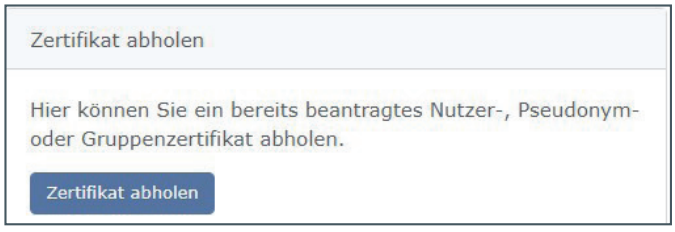


Abb. 17

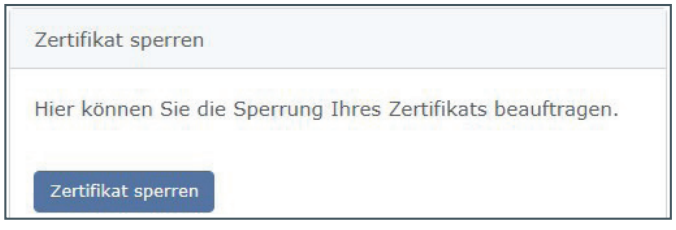


Abb. 18

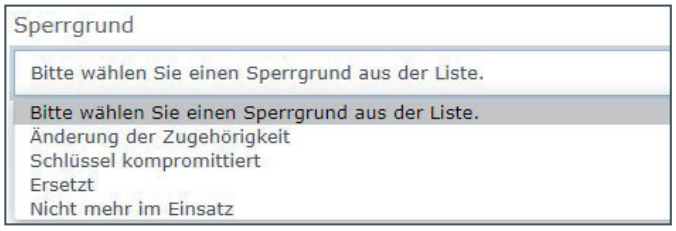


Abb. 20

Sehr geehrte Nutzerin, sehr geehrter Nutzer,
die Bearbeitung Ihres Zertifizierungsantrags 78243104 mit Profil User ist nun abgeschlossen.
Ihr Zertifikat mit der Seriennummer
112692116660281712752104727l9 ist auf den Namen
CN=Thorsten Hindermann,GN=Thorsten,SN=Hindermann,OU=AG O,O=Gesellschaft fuer wissenschaftliche Datenverarbeitung,L=GOETTINGEN,ST=NIEDERSACHSEN,C=DE
erstellt worden.
…

Abb. 19

den folgenden Artikeln der mehrteiligen Artikelserie „E-Mail-Verschlüsselung mit X.509-Zertifikaten“ nachgelesen werden:

- „Teil 2: Installation und Verteilung von Zertifikaten“ in den GWDG-Nachrichten 1-2/2020 [8]
- „Teil 3: Outlook-E-Mail-Anwendungen“ in den GWDG-Nachrichten 3/2020 [9]
- „Teil 4: Apple E-Mail-Anwendungen“ in den GWDG-Nachrichten 7-8/2020 [10]
- „Teil 5: Thunderbird, Notes und Mutt“ in den GWDG-Nachrichten 11/2020 [11]

Abb. 15

LINKS

[1] *https://info.gwdg.de/docs/doku.php?id=de:services:it_security:pki:start*

[2] *https://info.gwdg.de/docs/doku.php?id=de:services:it_security:pki: start#zertifikat_beantragen*

[3] *https://www.gwdg.de/documents/20182/27257/GN_12-2019_www.pdf#page=10*

[4] *https://www.gwdg.de/documents/20182/27257/GN_12-2020_www.pdf#page=11*

[5] *https://www.gwdg.de/documents/20182/27257/GN_4-5-2020_www.pdf#page=28*

[6] *https://www.gwdg.de/documents/20182/27257/GN_12-2020_www.pdf#page=17*

[7] *https://www.gwdg.de/documents/20182/27257/GN_3-2021_www.pdf#page=11*

[8] *https://www.gwdg.de/documents/20182/27257/GN_1-2-2020_www.pdf#page=14*

[9] *https://www.gwdg.de/documents/20182/27257/GN_3-2020_www.pdf#page=6*

[10] *https://www.gwdg.de/documents/20182/27257/GN_7-8-2020_www.pdf#page=8*

[11] *https://www.gwdg.de/documents/20182/27257/GN_11-2020_www.pdf#page=12*

Änderung der Zugehörigkeit
Dieser Sperrgrund signalisiert, dass der Zertifikatname (SubjectDN) oder andere Informationen im Zertifikat nicht mehr den aktuellen Tatsachen entsprechen, etwa weil sich Namen, E-Mail-Adressen, Funktionsrollen oder Zugehörigkeiten geändert haben, dass aber gleichzeitig kein Grund für einen Verdacht besteht, dass der zum Zertifikat gehörende geheime Schlüssel kompromittiert (offengelegt) wurde.

Schlüssel kompromittiert
Dieser Sperrgrund signalisiert, dass der zum Zertifikat gehörende geheime Schlüssel kompromittiert (offengelegt) wurde oder andere Aspekte der durch den Zertifikatnamen (SubjectDN) beschriebenen Entität kompromittiert wurden. Dieser Sperrgrund ist ebenfalls zu wählen, wenn der bloße Verdacht einer solchen Kompromittierung besteht.

Ersetzt
Dieser Sperrgrund signalisiert, dass das Zertifikat von einem neueren Zertifikat abgelöst wurde, dass aber gleichzeitig kein Grund für einen Verdacht besteht, dass der zum zu sperrenden Zertifikat gehörende geheime Schlüssel kompromittiert (offengelegt) wurde.

Nicht mehr im Einsatz
Dieser Sperrgrund signalisiert, dass das Zertifikat nicht länger für die Zwecke eingesetzt wird, für die es ausgestellt wurde, dass aber gleichzeitig kein Grund für einen Verdacht besteht, dass der zum Zertifikat gehörende geheime Schlüssel kompromittiert (offengelegt) wurde.

Abb. 21

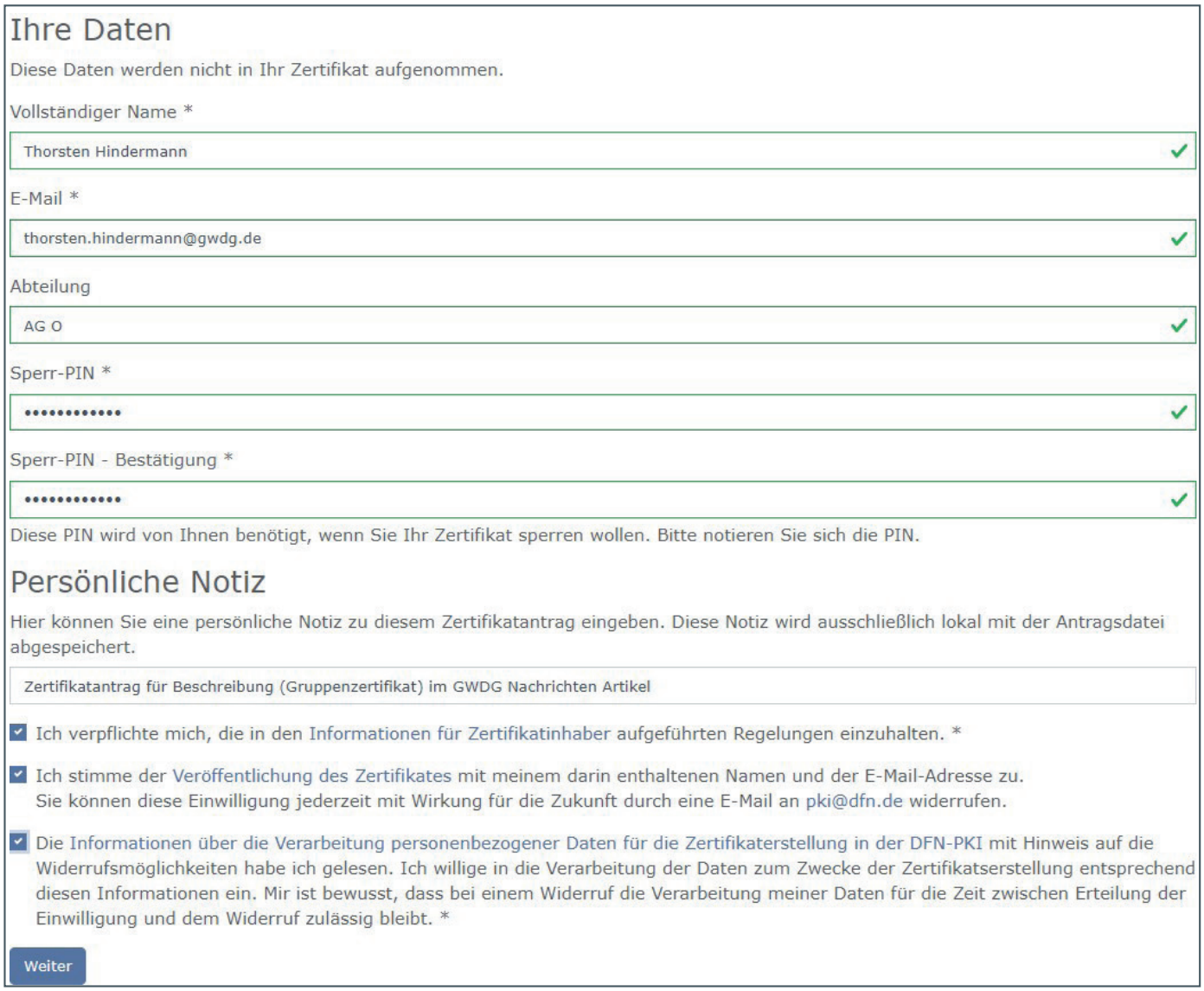

Ihre Daten
Diese Daten werden nicht in Ihr Zertifikat aufgenommen.
Vollständiger Name *
Thorsten Hindermann
E-Mail *
thorsten.hindermann@gwdg.de
Abteilung
AG O
Sperr-PIN *
Sperr-PIN - Bestätigung *
Diese PIN wird von Ihnen benötigt, wenn Sie Ihr Zertifikat sperren wollen. Bitte notieren Sie sich die PIN.
Persönliche Notiz
Hier können Sie eine persönliche Notiz zu diesem Zertifikatantrag eingeben. Diese Notiz wird ausschließlich lokal mit der Antragsdatei abgespeichert.
Zertifikatantrag für Beschreibung (Gruppenzertifikat) im GWDG Nachrichten Artikel
Ich verpflichte mich, die in den Informationen für Zertifikatinhaber aufgeführten Regelungen einzuhalten. *
Ich stimme der Veröffentlichung des Zertifikates mit meinem darin enthaltenen Namen und der E-Mail-Adresse zu.
Sie können diese Einwilligung jederzeit mit Wirkung für die Zukunft durch eine E-Mail an pki@dfn.de widerrufen.
Die Informationen über die Verarbeitung personenbezogener Daten für die Zertifikaterstellung in der DFN-PKI mit Hinweis auf die Widerrufsmöglichkeiten habe ich gelesen. Ich willige in die Verarbeitung der Daten zum Zwecke der Zertifikatserstellung entsprechend diesen Informationen ein. Mir ist bewusst, dass bei einem Widerruf die Verarbeitung meiner Daten für die Zeit zwischen Erteilung der Einwilligung und dem Widerruf zulässig bleibt. *
Weiter

Abb. 16

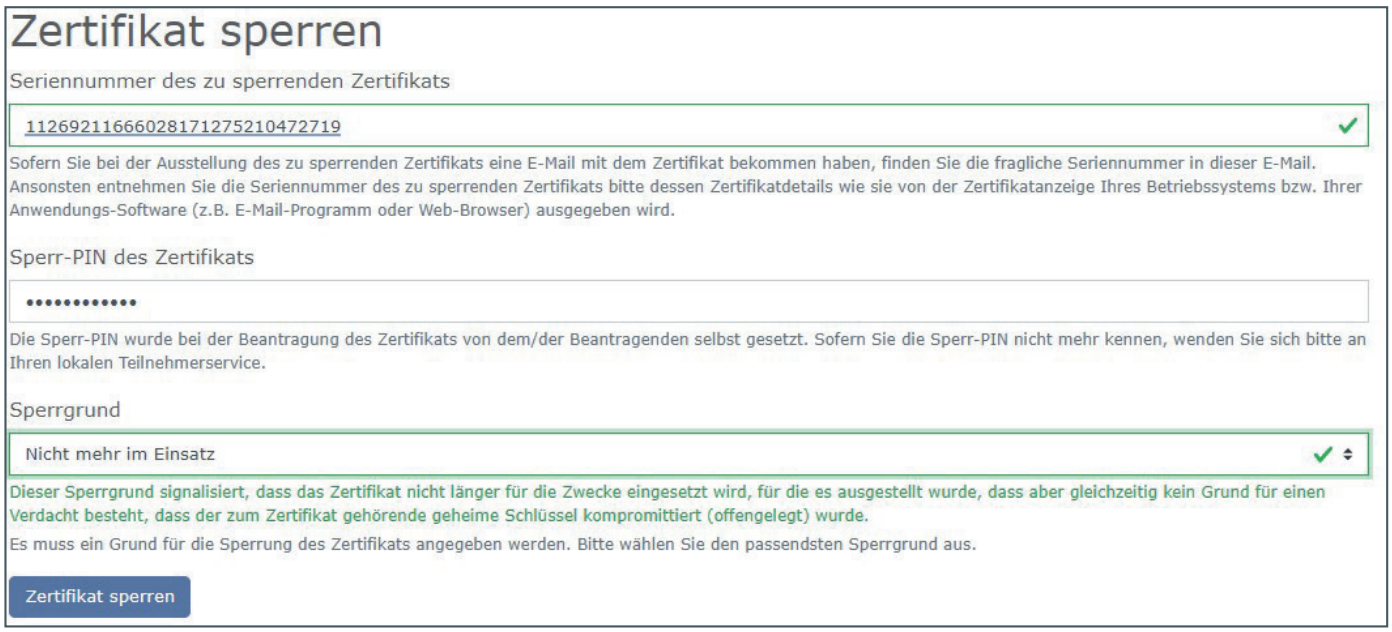

Zertifikat sperren
Seriennummer des zu sperrenden Zertifikats
112692116660281712752104727 19
Sofern Sie bei der Ausstellung des zu sperrenden Zertifikats eine E-Mail mit dem Zertifikat bekommen haben, finden Sie die fragliche Seriennummer in dieser E-Mail. Ansonsten entnehmen Sie die Seriennummer des zu sperrenden Zertifikats bitte dessen Zertifikatdetails wie sie von der Zertifikatanzeige Ihres Betriebssystems bzw. Ihrer Anwendungs-Software (z.B. E-Mail-Programm oder Web-Browser) ausgegeben wird.
Sperr-PIN des Zertifikats
Die Sperr-PIN wurde bei der Beantragung des Zertifikats von dem/der Beantragenden selbst gesetzt. Sofern Sie die Sperr-PIN nicht mehr kennen, wenden Sie sich bitte an Ihren lokalen Teilnehmerservice.
Sperrgrund
Nicht mehr im Einsatz
Dieser Sperrgrund signalisiert, dass das Zertifikat nicht länger für die Zwecke eingesetzt wird, für die es ausgestellt wurde, dass aber gleichzeitig kein Grund für einen Verdacht besteht, dass der zum Zertifikat gehörende geheime Schlüssel kompromittiert (offengelegt) wurde.
Es muss ein Grund für die Sperrung des Zertifikats angegeben werden. Bitte wählen Sie den passendsten Sperrgrund aus.
Zertifikat sperren

Abb. 22

Kubernetes with Rancher at the GWDG – Part 2: Setup and Deployment

Text and Contact:
Samaneh Sadegh
samaneh.sadegh@gwdg.de
0551 201-2113

Rancher is a popular open source tool to deploy and manage Kubernetes clusters. In this second part of a series of articles, we explain how to install Rancher on premise and use its web interface to deploy a Kubernetes cluster. Then we briefly introduce the command-line tools which can be used to interact with deployed clusters. Future articles will then show how to use Rancher to deploy different applications into the managed clusters.

INTRODUCTION

In part 1 (see the GWDG News 3/2021) of this series of articles a short introduction to Kubernetes as a management framework for container-based applications and its architecture was given. In this article we will concentrate on the installation of Rancher as a tool to easily deploy and manage Kubernetes clusters on premise.

INSTALLING RANCHER

As shown in figure 1 there are several ways to install Rancher which should be chosen based on individual requirements.

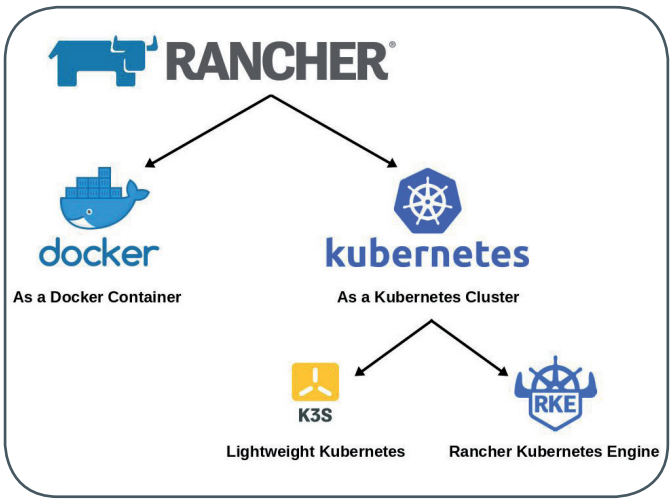


Figure 1: Different Rancher deployment options

For testing or demonstration purposes, you can install Rancher in a single *Docker container*. In a Docker installation, you can explore the Rancher functionality without using too much resources. However, this deployment option is suitable only for development and testing purposes.

For production environments, it is recommended to install Rancher in a Kubernetes cluster itself allowing for scalability and redundancy. For this, you need to provide a Kubernetes cluster first which must be based either on the *K3s* or the *RKE Kubernetes distributions*. However, a K3s cluster is recommended as it uses far less resources (the binary is less than 100 MB). Besides, it allows an external datastore to hold the cluster data, allowing the K3s server nodes to be treated as ephemeral. In the following, a Rancher install on a highly-available K3s cluster will be explained. For other installation types, please check the official website of Rancher (*https://rancher.com/*).

To install Rancher on a HA K3s cluster, the following four steps for preperation are required (for more details please check Ranchers official website *https://rancher.com/docs/k3s/latest/en/installation/ha/*):

1. Deploy two virtual machines (VMs) with enough resources (e.g. 8 GB RAM, 2 CPUs and 20 GB HDD with Ubuntu 20.04 LTS) where the components of Rancher will be running later. Then carefully setup firewall rules based on Ranchers official document (*https://rancher.com/docs/rke/latest/en/os/#ports*).
2. Setup a load balancer (e.g. HAProxy) for both ports 80 and 443 for the two VMs. Then setup a DNS entry for your load balancers IP.
3. Setup a SQL database or use an already existing one. As

Kubernetes mit Rancher bei der GWDG – Teil 2: Einrichtung und Bereitstellung

Rancher ist ein beliebtes Open-Source-Tool zur Bereitstellung und Verwaltung von Kubernetes-Clustern. In diesem zweiten Teil einer Artikelserie erklären wir, wie Sie Rancher lokal installieren und danach einen Kubernetes-Cluster über dessen Web-Oberfläche ausrollen können. Anschließend stellen wir kurz einige wichtige Kommandozeilen-Tools vor, die zur Interaktion mit den installierten Clustern verwendet werden können. Zukünftige Artikel werden dann zeigen, wie mit Hilfe von Rancher container-basierte Anwendungen in die verwalteten Kubernetes-Cluster installiert werden können.

a simple way for a test environment, you can deploy e.g. MariaDB in a separate VM. At the GWDG, a high-availability Galera cluster is used for redundancy and scalability. Then create a database along with a user and password for K3s with an "all" grant on the database.

4. Finally, use "ssh" to connect to both VM instances provided for Rancher and configure them by running the following command adapted to your database setup:

```
curl -sfL https://get.k3s.io | sh -s - server --data-store-endpoint="mysql://username:password@tcp(hostname:3306)/database-name"
```

After providing the cluster, Rancher along with its required components (e.g. *cert-manager*) can be installed using *Helm* as follows:

Use "ssh" to connect to one of the Rancher nodes and install Helm:

```
sudo snap install helm --classic
```

Add required repositories to Helm:

```
helm repo add rancher-stable https://releases.rancher.com/server-charts/stable
helm repo update
helm repo add jetstack https://charts.jetstack.io
helm repo update
```

Then create two Kubernetes namespaces for deploying Rancher and cert-manager:

```
kubectl create namespace cattle-system
kubectl create namespace cert-manager
```

There are some resources required by cert-manager which should be installed separately:

```
kubectl apply –validate=false -f https://github.com/jetstack/cert-manager/releases/download/v1.0.4/cert-manager.crds.yaml
```

Then you can install cert-manager:

```
helm install cert-manager jetstack/cert-manager –kubeconfig /etc/rancher/k3s/k3s.yaml --namespace cert-manager --version v1.0.4
```

Wait for *Pods* to be running:

```
kubectl get pods --namespace cert-manager
```

Finally, install Rancher while providing the DNS address you setup for the load balancer:

```
helm install rancher rancher-stable/rancher –kubeconfig /etc/rancher/k3s/k3s.yaml --version v2.4.15 --namespace cattle-system –set hostname="your DNS address"
```

Check the deployment status and wait until it is finished:

```
kubectl -n cattle-system rollout status deploy/rancher
```

Now you can access Rancher dashboard using the DNS address you provided. At the first login, Rancher will ask you to set an admin password.

DEPLOYING KUBERNETES WITH RANCHER

In this section, we explain how to use Ranchers web interface to deploy a Kubernetes cluster. Although Rancher supports several ways to deploy a cluster, we explain two of them which are used at the GWDG.

Custom cluster

The custom cluster option is used when the required infrastructure resources shall be provided independent from Rancher. At the GWDG, *ESX VMs* are used to deploy a Kubernetes cluster in this mode.

To have a custom Kubernetes cluster managed by Rancher, four main steps should be followed:

1. Provide the infrastructure resources including virtual machines, networks, a load balancer and security policies.
2. Install Docker in all the VMs (the version should be compatible with the installed Rancher version).
3. Create a custom Kubernetes cluster using the Rancher UI (see figure 2).

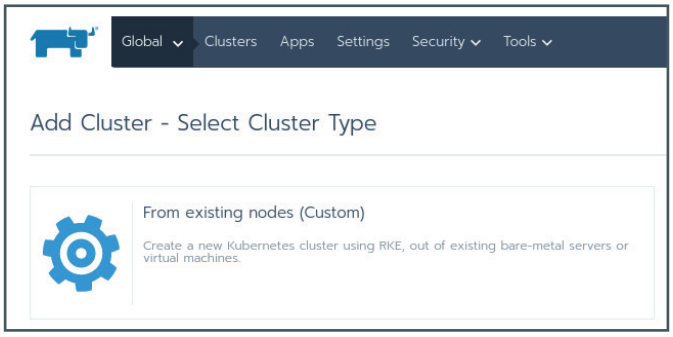


Figure 2: Rancher UI – Create a custom cluster

4. Run the script provided by Rancher in the previous steps on all the VMs according to their role (see figure 3).

At GWDG, these steps are automated by configuration management tools (*Puppet* and *Foreman*) which helps to make provisioning a Kubernetes cluster streamlined and reliable.

OpenStack as cloud provider

Rancher is able to create the required VMs itself by directly communicating with the APIs of an IaaS platform (called "cloud providers" in Rancher). At the GWDG, *OpenStack* can be used to deploy a Kubernetes cluster in this fashion, which we will describe in the following as an example on how to use this method. Although Rancher also supports *VMware vSphere* as a "cloud provider" it is currently not possible to use the VMware vSphere environment at GWDG in this way due to security considerations.

To use OpenStack as a cloud provider, first you should enable OpenStack in the Rancher UI, and then provide Rancher with the required API endpoints and access. To do this, five Steps should be followed:

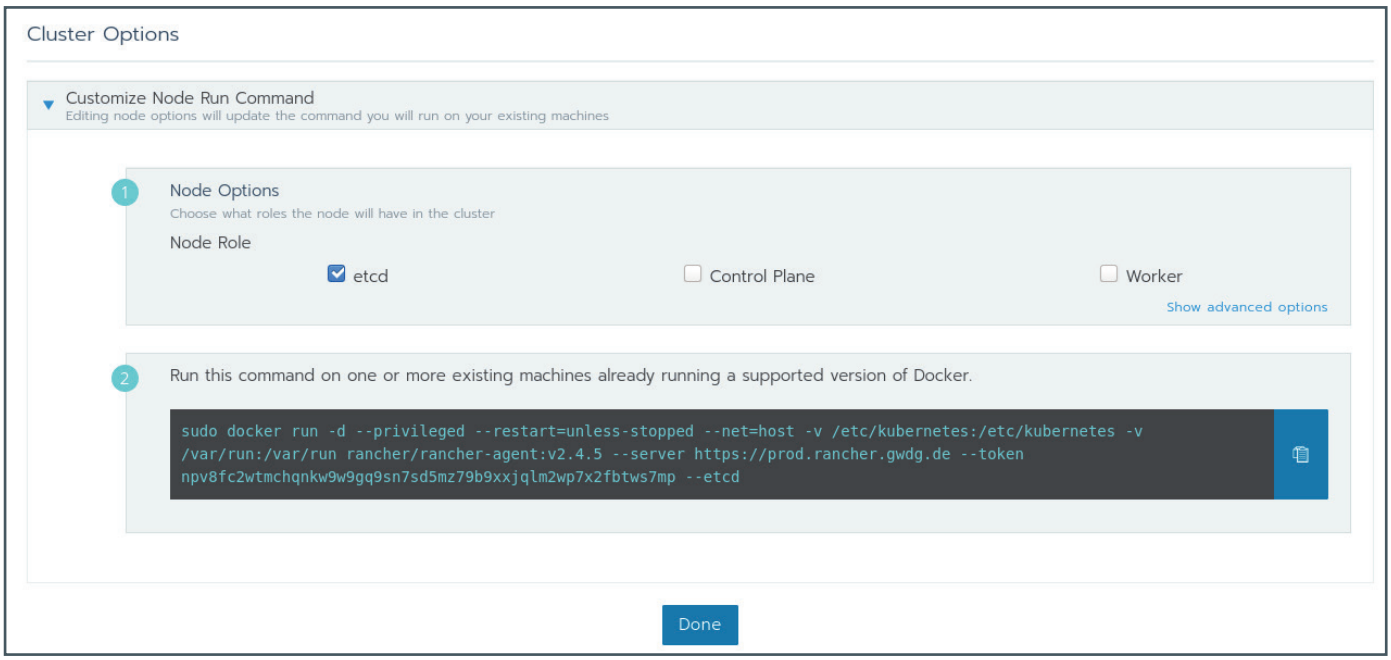


Figure 3: Rancher UI – Run the custom cluster final script

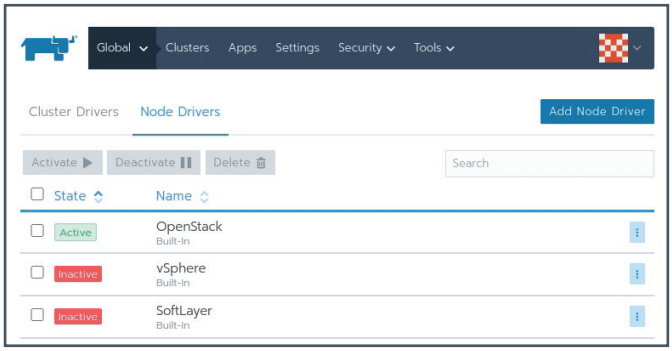


Figure 4: Rancher UI – Activate the OpenStack node driver

1. Enable the OpenStack node driver using Rancher UI. First select "Tools" > "Drivers", and then from the Node Drivers tab, enable OpenStack (see figure 4).
2. Define a node template by selecting "Node Templates" from the small menu under the profile picture on the top-right of the Rancher UI (see figure 5). Then click "Add Template" and fill out the displayed form based on your account data in OpenStack. Notice that one of the required information is a security group name (the field "secGroups") which should be pre-defined in your OpenStack account taking the Kubernetes port requirements into consideration.

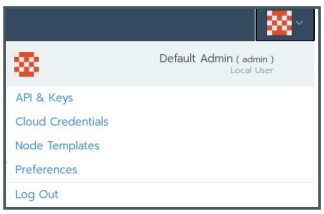


Figure 5: Rancher UI – Small menu under profile picture

3. Enable automatic storage provisioning by activating additional storage-drivers in "Advanced Settings" > "Feature Flags" (see figure 6).
4. Create the Kubernetes cluster by clicking "Add Cluster" at the "Global" level and selecting OpenStack as provider. In the displayed form, define a name for the cluster, a prefix for node names, select the node template, set each node role and number (see figure 7) and finally click "Edit as YAML". Add a "cloud_provider" section to the displayed "YAML" based on the following form providing your OpenStack account information:

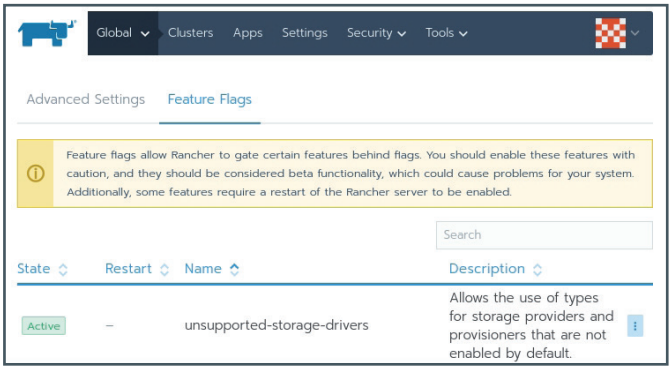


Figure 6: Rancher UI – Activate additional storage-drivers

```
cloud_provider:
  name: openstack
  openstackCloudProvider:
    global:
      username: <yourUserName>
      password: <yourPassword>
      auth-url: <Keyston_public_url>
      tenant-id: <yourProjectID>
      domain-name: GWDG
    load_balancer:
      create-monitor: false
      floating-network-id: <yourPublicNetworkID>
      manage-security-groups: true
      monitor-max-retries: 0
      subnet-id: <yourSubnetID>
      use-octavia: false
    block_storage:
      ignore-volume-az: true
    route:
      router-id: <yourRouterId>
```

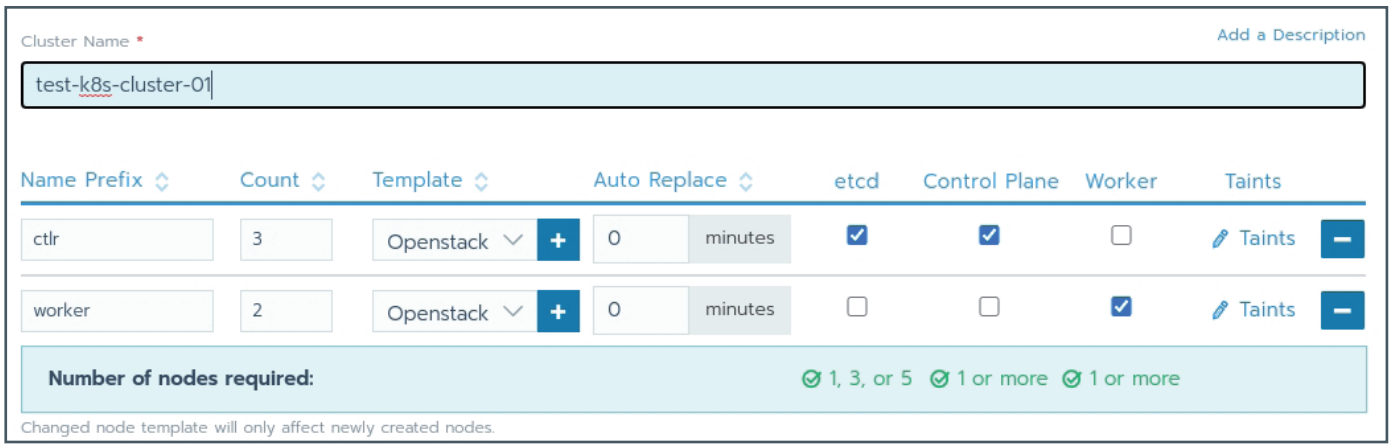


Figure 7: Rancher UI – Configuration for the cluster nodes

5. Configure persistence volumes for your cluster by defining suitable storage classes at "Storages" > "Storage Classes".

After completing all these steps, you will be able to see the automatic creation of the VMs and pair keys in your OpenStack dashboard. At the GWDG, a mechanism to automate all the mentioned steps is under development, so that it will become very easy to roll out Kubernetes clusters under OpenStack accounts which can be then managed through the Rancher UI.

CONNECT TO YOUR CLUSTER USING A CLI

In this section, we introduce two command line tools to interact with your new Kubernetes cluster: *Kubernetes native CLI* (which was already used in the first section to install Rancher) and *Rancher CLI*.

The Kubernetes command-line tool, *kubectl*, can be installed on your local machine and used to communicate with your remote cluster using a config file. To provide the config file, create a local YAML file (e.g. *cluster.yaml*) and set its contents from the Rancher UI (see figure 8) on the cluster page.

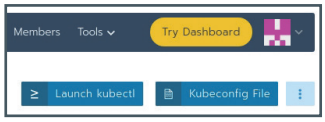


Figure 8: Rancher UI – "Kubeconfig File" to connect to a Kubernetes cluster via CLI

This config file should be either copied to *~/.kube/config* or provided to the *kubectl* command as a flag in the way we are using in the following examples in this section.

To start, list the nodes in your cluster to check if you are using the right config file. This command displays a list of the VMs you have used to deploy the Kubernetes cluster:

```
kubectl --kubeconfig cluster.yaml get nodes
```

Then, list the namespaces to select the one you intend to use for further queries:

```
kubectl --kubeconfig cluster.yaml get ns
```

List the pods inside your desired namespace (e.g. "cattle-system") to check the state of the running services:

```
kubectl --kubeconfig cluster.yaml get pods -n <ns>
```

Check the logs of your running application by specifying its related pod:

```
kubectl --kubeconfig cluster.yaml logs -n <ns> <pod>
```

If the pod owns more than one container, kubectl would show an error along with a list of the containers to select from. In this case, the logs of a specific container should be queried as follows:

```
kubectl --kubeconfig cluster.yaml logs -n <ns> <pod>
<container>
```

To have a quick help on possible commands, run *kubectl --help* and to check the list of all available commands, look at *https://kubernetes.io/docs/reference/generated/kubectl/kubectl-commands*.

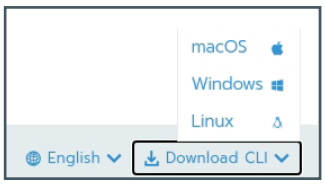


Figure 9: Rancher UI – Download the Rancher CLI from bottom-right of any page

The Rancher command-line tool can be downloaded from the Rancher UI on the bottom-right of any page (see figure 9). Then, copy the executable file in one of your *$PATH* directories (e.g. */usr/local/bin/*).

To connect to your Kubernetes cluster using the Rancher CLI, you need an access key which can be generated from the "API & Keys" form on the small menu under the profile icon (see figure 5). Afterwards, run the following command and select the desired project from the list displayed by the Rancher CLI:

```
rancher login <rancher-endpoint> --token user:pass
```

The Rancher CLI will create a *json* file containing the required information in */Users/<user>/.rancher/* (here MacOS X was used). Thus, as long as you are using the same local machine and user, you don't need to login again. To change the selected project run:

```
rancher context switch
```

To check the current selected project run:

```
rancher context current
```

The command *rancher kubectl* is an alternative for the native *kubectl* command which does not need the config file as Rancher

CLI already has all the required information inside its own *json* file. Therefore, you can run the following command to get the logs of an application running inside the selected project:

```
rancher kubectl logs -n <ns> <pod> <container>
```

Run *rancher help* to have a view of possible commands or visit *https://rancher.com/docs/rancher/v2.x/en/cli* for more information.

In the next part of these series of articles we will show how to deploy container-based applications using the Rancher UI into the deployed/managed Kubernetes cluster.

ePIC
Persistent Identifiers for eResearch

Stable References for Research Data

Text and Contact:
Dr. Sven Bingert
sven.bingert@gwdg.de
0551 201-2164

The GWDG, as partner of ePIC [1] and Multi-Primary Administrator in DONA, offers Persistent Identifier (PID) services for users, communities and institutions. These services include PID self-services, the registration of namespaces (so called prefixes) and consulting to find the best solution to integrate PIDs in the research data management. Additionally, the eScience group of the GWDG continuously investigates new ideas, develops and implements new additional PID services.

A perfect Research Data Management (RDM) seems to become the holy grail at all levels. Organizing research data as a PhD student up to entire institutes, is a challenging task with many objectives. These developments are encouraged and sometimes driven by international communities defining standards and processes for the RDM. E.g. the FAIR principles [2] list attributes for research data while the Digital Object Architecture [3] focuses more on the technical implementations. All these proposals follow the goal to foster archiving, sharing, and reusing of the valuable research data. One thing they have in common is the requirement to assign an (persistent) identifier to each research data object. Persistent Identifier (PID) is not a new concept and many implementations for PID systems are in use (see e.g. [4], [5] or [6]). The Handle system [7] is well known and is the technical base of PID providers such as the DOI Foundation and ePIC.

EPIC PID SERVICE

The GWDG offers Persistent Identifier services based on the Handle system. These services are aimed at different users, groups, or institutes based on their requirements. The GWDG also offers consulting to identify and to define the requirements before the service will be initiated. Using the GWDG PID services a single PID can be created (or minted) using the self-service in the GWDG customer portal [8] but also massive amount of PIDs in assigned namespaces (so called prefix) can be created. The offered service is as follows:

1. Creation of single PIDs with a limited set of metadata using the GWDG customer portal. These PIDs are assigned to the logged-in user. Using the self-service, the PID content (e.g. the URL) can be modified if the location of the referenced object changes.
2. Requesting a namespace for an institution or community. The technical infrastructure of the GWDG will be used to create and store PIDs while the data objects will be maintained by the institution.
3. Requesting a namespace as in 2. but also setting up the necessary technical PID infrastructure within the requesting institution.

The main difference between the two last service options is, that in 2. the PID data base will be mirrored within the ePIC partners [1] as part of the ePIC Quality of Service and Policies [9]. This defines a high-availability service for Persistent Identifiers. For

Stabile Referenzen für Forschungsdaten

Die GWDG bietet etablierte Dienste zur Erstellung und Verwaltung von Persistenten Identifikatoren (PIDs). Die Dienste basieren auf der Handle-Software [7] und sind durch die Partnerschaft in ePIC [1] hochverfügbar. Neben der Bereitstellung der technischen Infrastruktur und Einrichtung neuer Namensräume (Präfixe) unterstützt die GWDG durch Beratung zur Anforderungsanalyse bei Forschungsdatenmanagement-Prozessen. Zusätzlich werden in eigenen Forschungsprojekten und Abschlussarbeiten PID-Daten untersucht und neue Dienste entwickelt. Dabei werden internationale Entwicklungen zur PID-Standardisierung berücksichtigt und Dienste übernommen. Als Beispiel ist hier die Data Type Registry (DTR) zu nennen, die zur Standardisierung und Validierung von PID-Datenfeldern (Types) genutzt werden kann.

creating and maintaining PID data (e.g. URL or OWNER) a REST-API is available. This allows to integrate and automate the PID maintenance in data management workflows.

As part of our research department, we are interested to build new services on top of the basic PID service. One important topic is the quality of the given PID data. The PID data is structured in key-value pairs, e.g., the key is *URL* and the value is *www.gwdg.de/dataobject*. In the PID world the key is also called type. Using the PID service each creator has, in principle, the possibility to define a new type. But that leads to highly heterogenous types and the reuse of the PID data and the referenced data objects is aggravated. A simple example is the type of the year a data object was published. That could be *pubyear*, *pub-year*, *publication-year*, and so on. In order to reach a certain level of standardization the GWDG operates a Data Type Registry (DTR) [10] where types can be defined and a validation mechanism placed. As the DTR type definitions are also based on PIDs, the use of types and the evaluation of PID data can be integrated in machine actionable workflows.

PID RESEARCH

In a research project (bachelor thesis "Analysis of a large database of persistent identifiers using graph technologies" by Elly Schmidt) we investigate the PID data of different namespace reflecting different communities or institutions. In this project we pushed the PID data of several millions of PIDs in a graph data base. The goal was to identify similarities between namespaces or issues in the given PID data. Figure 1 depicts the schema used when building the graph. We can identify types which are defined in the DTR (hierarchical) and types which are only used as key-value pairs (flat).

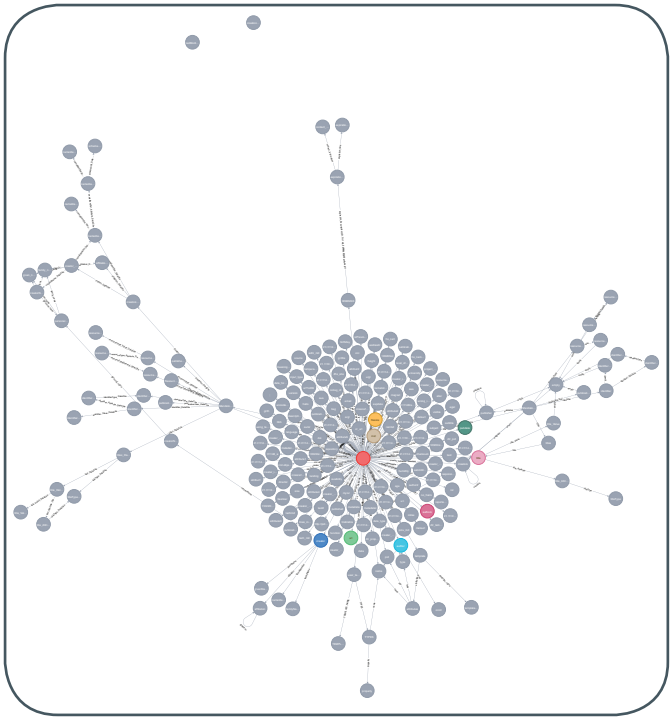


Figure 1: Graph schema used for the PID data. Highlighted with color are types (nodes) which are most frequent in the data base (e.g. url, inst). The red central node is the PID itself. The figure highlights the structure of the schema, while single node lables are not of importance.

The graph analysis showed that the number of types used per PID is still very low (3.5 in average) and the DTR, as it is a relatively new concept, is not used intensively. Other interesting graph related results will be soon available.

The most important type of the PID is *URL*, which is the actual reference to the data object. It is the responsibility of the data provider (sometimes equal to data owner) to update this type if the location of the object changes. That could be in the case if the domain for a repository changes or a website is restructured. Sometimes the PID creator is not the data provider, e.g., by referencing to a third-party URL. In that case the PID creator might not be aware of changes and therefore the necessary update of the PID data get lost. To support the PID creator and data provider we implemented a simple service testing PIDs and collect that information. The results can then be evaluated by the data provider and PID updates, if necessary, undertaken.

Figure 2 depicts an example output for 23 different namespaces. The majority of the URLs return with status code *ok*. But there is still a large number of URLs which return either *not found* or *unreachable hosts*. There are different reasons to explain these numbers. E.g., a PID could be created before the data object is (publicly) available and the URL type would be updated later. Therefore, only the data provider (and owner of the namespace) is capable to understand the given statistics.

In Figure 3 we can see that in three prefixes almost 100% of the tested PIDs return with a status code *ok*. But there is also a prefix where over 90% of the tests return a *internal server error*. Of course the total number of tests for each prefix is required to evaluate these numbers correctly.

SUMMARY

Besides the established PID services, the GWDG plans and develops new services to improve the usage for research data. We follow international developments in PID standardization and adapt services such as the DTR in our service portfolio. In our presentations and collaborations with partners and users we identify requirements such as smart search interfaces or applying standards to the PID data.

LINKS

[1] *https://www.pidconsortium.net*
[2] *https://www.go-fair.org*
[3] *https://www.dona.net/digitalobjectarchitecture*
[4] *https://transportation.libguides.com/persistent_identifiers*
[5] Hakala, J. (2010): Persistent identifiers – an overview. Retrieved from *http://www.metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/*
[6] Plomp, E. (202): Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments. In: Data Science Journal, 19 (1), p. 46., DOI: *http://doi.org/10.5334/dsj-2020-046*
[7] *https://www.handle.net*
[8] *https://www.gwdg.de/application-services/persistent-identifier-pid*
[9] *http://www.pidconsortium.net/?page_id=904*
[10] *http://dtr-pit.pidconsortium.net*

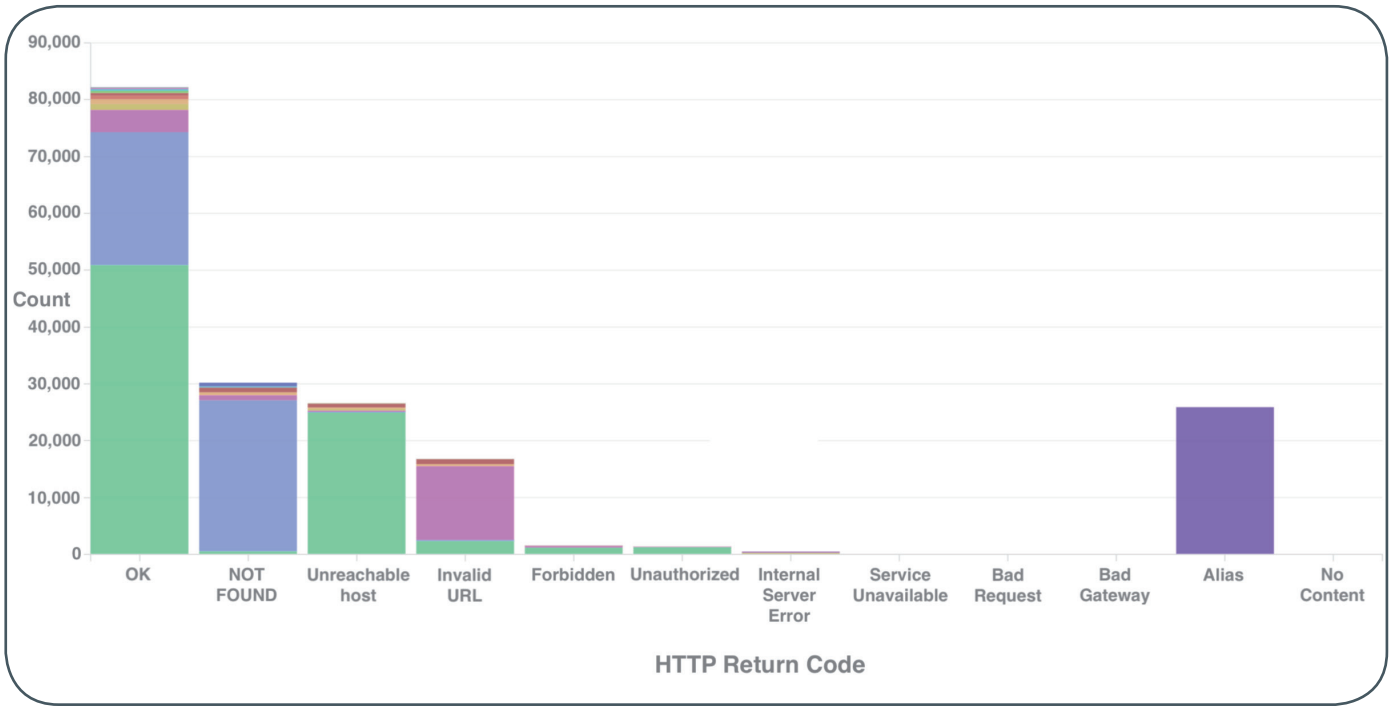


Figure 2: HTTP status codes for tested PIDs. Color coded are the different prefixes (namespaces). The value 'Alias' is not a status code but is used for PIDs which are aliases for other PIDs.

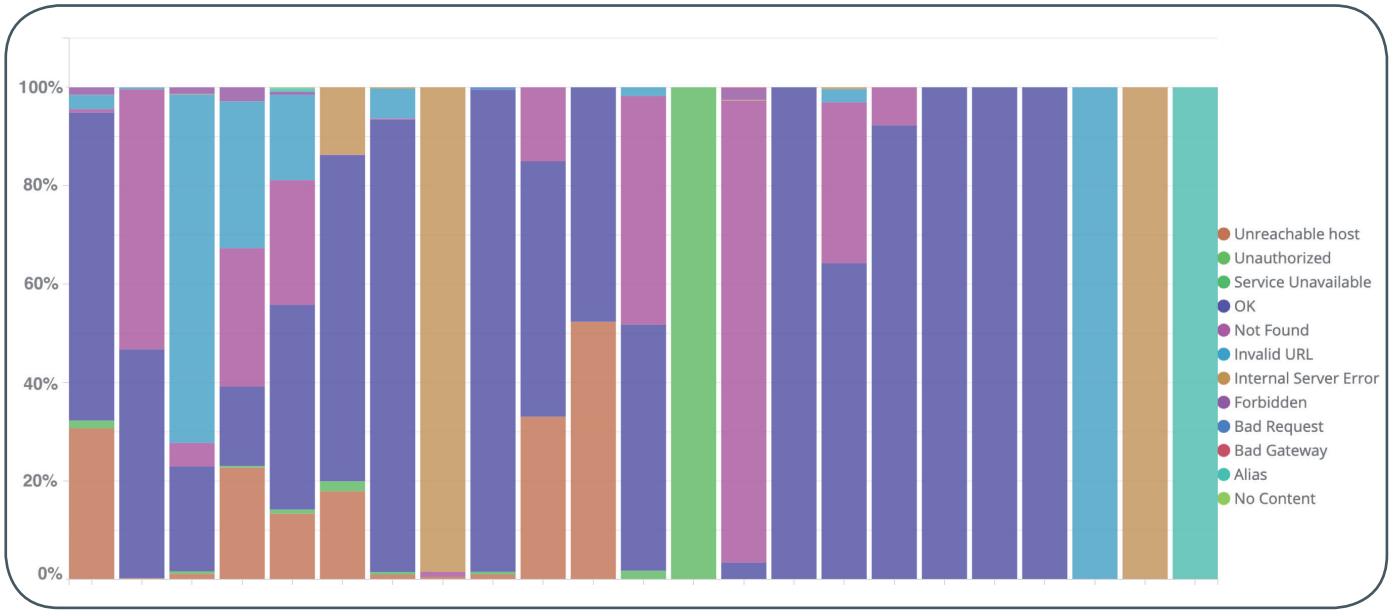


Figure 3: Relative number of HTTP status code replies (color coded) per prefix. Each column represents a prefix (23 in total).

Stellenangebot
Nr. 20210309

Die GWDG sucht zum nächstmöglichen Zeitpunkt zur Verstärkung des High-Performance-Computing-Teams der Arbeitsgruppe „eScience“ (AG E)

HPC-Expert*innen (m/w/d) mit Erfahrungsschwerpunkten in einem oder mehreren der Bereiche Anwendungen, Betrieb und Methodik

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 Ent-geltgruppe TVöD E 13, bei besonderer Eignung bis TVöD E14 vorgesehen. Die Stellen sind zunächst befristet. Die GWDG strebt eine langfristige Zusammenarbeit an. Bei Interesse besteht die Mög-lichkeit zur Promotion.

Themengebiet

Zur Verstärkung unseres High-Performance-Computing-Teams suchen wir engagierte Mitarbeiter*innen mit einem nachgewiesenen Interesse an den Herausforderungen des Hochleistungsrechnens. Sie möchten an der Weiterentwicklung unseres HPC-Standortes und des NHR-Verbundes mitwirken, die Performance wissenschaftlicher Anwendungen und Systeme optimieren oder neue Forschungsthemen im Göttinger HPC-Umfeld etablieren? Dann bewerben Sie sich!

Aufgabenbereiche

- Beratung zur effizienten Nutzung der verfügbaren Rechen- und Speicherressourcen
- Durchführung eigener Forschungsarbeiten
- Mitarbeit bei der Administration und kontinuierlichen Weiterentwicklung der HPC-Dienste und deren Infrastruktur
- Durchführung von Fehler- und Performance-Analysen von wissenschaftlichen Anwendungen und Mitwirkung an deren Optimierung
- Analyse der Systemnutzung zur Erkennung von Optimierungs- und Entwicklungspotenzialen der HPC-Systeme
- Regelmäßige Mitwirkung bei Workshops und Schulungen

Anforderungen

- Abgeschlossenes Hochschulstudium oder vergleichbare Qualifikation mit einschlägiger Berufserfahrung
- Gutes analytisches Denkvermögen
- Selbstständige, strukturierte und systematische Arbeitsweise
- Ausgeprägte Team- und Kommunikationsfähigkeit
- Sehr gute Deutsch- und Englischkenntnisse in Wort und Schrift

Aufgrund der Schwerpunktbildung unseres Zentrums im NHR-Verbund suchen wir insbesondere Kandidat*innen mit Erfahrung in einem oder mehreren der folgenden Gebiete:

- HPC-Nutzung in Lebenswissenschaften, Digital Humanities, Erdsystemwissenschaften und Numerische Strömungsmechanik
- Virtualisierung, Containerlösungen, Softwarebereitstellung, Big Data und KI-Systeme, IO und Speichersysteme für HPC
- Administration von HPC-Clustern oder anderen Linux-Server-Farmen und effiziente Nutzung von HPC-Systemen
- Management von wissenschaftlichen oder Entwickler-Communities

Unser Angebot

- Flexible Arbeitszeiten und Möglichkeit zu mobilem Arbeiten
- Ein modernes, vielfältiges und außergewöhnliches Arbeitsumfeld mit großer Nähe zu Wissenschaft und Forschung an der Schnittstelle mehrerer innovativer Technologiesektoren
- Eine interessante, vielseitige Tätigkeit in einem großen, international agierenden IT-Kompetenzzentrum
- Mitarbeit in einem kompetenten und engagierten Team
- Qualifizierung und Weiterentwicklung Ihrer Fähigkeiten
- Im öffentlichen Dienst übliche Sozialleistungen, wie z. B. Betriebsrente mit zusätzlicher Erwerbsminderungs- und Hinterbliebenenversorgung

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 über unser Online-Formular unter *https://s-lotus.gwdg.de/gwdgdb/age/20210309.nsf/bewerbung*.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

Herr Dr. Christian Boehme
Tel.: 0551 201-1839
E-Mail: *christian.boehme@gwdg.de* oder das

HPC-Team
E-Mail: *hpc-team@gwdg.de*

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Stellenangebot
Nr. 20210304

Bei der GWDG ist zum nächstmöglichen Zeitpunkt zur Verstärkung des Identity- und Access-Management-Teams der Arbeitsgruppe „Basisdienste und Organisation“ (AG O) eine Stelle als

Softwareentwickler*in (m/w/d) im Bereich Identity- und Access-Management

zu besetzen.
Die regelmäßige Wochenarbeitszeit beträgt 39 Stunden. Die Stelle ist zunächst auf 24 Monate befristet; eine Option zur unbefristeten Übernahme besteht. Die Vergütung erfolgt nach dem Tarifvertrag für den öffentlichen Dienst (Bund); die Eingruppierung ist je nach Qualifikation bis zur Entgelt-gruppe TVöD E 13 vorgesehen.

Themengebiet
Der Bereich Identity- und Access-Management befasst sich mit der Automatisierung von Prozessen, die notwendig sind, um die Zugänge und Berechtigungen zu Systemen zu erhalten. Ihre Aufgabe wird es sein, die bestehende Authentifizierungs- und Autorisierungs-Infrastruktur zu pflegen und weiterzuentwickeln. Weiterhin zählen Projekte rund um das Thema Identity- und Access-Management wie z. B. Mehr-Faktor-Authentifizierung zu Ihrem Aufgabenbereich. Zusätzlich zählt zu Ihrem Themengebiet Softwareentwicklung in der Programmiersprache PHP.

Aufgabenbereiche
- Technische Konzeption, Implementierung und Dokumentation von Prozessen
- Weiterentwicklung von Anwendungen in PHP und anderen Programmiersprachen
- Anbindung von neuen Systemen und Applikationen an die bestehende Authentifizierungs- und Autorisierungs-Infrastruktur
- Wartung und Betrieb von Linux-Servern
- Durchführung von Softwaretests und Unterstützung der produktiven Inbetriebnahme der entwickelten Lösungen
- Unterstützung des 2nd/3rd-Level-Supports

Anforderungen
- Bachelor- oder Masterabschluss im Bereich Informatik oder eine vergleichbare Qualifikation und mindestens zwei Jahre Berufserfahrung im Bereich der Softwareentwicklung
- Fundierte Kenntnisse in einer oder mehreren Programmiersprachen wie PHP, Java EE oder C#/.NET
- SAML und OpenID Connect sind für Sie keine Fremdwörter.
- Gute Kenntnisse im Umgang mit Linux-Servern
- Spaß an einem freundlichen und lösungsorientierten Umgang mit Kund*innen und Arbeit in einem Team
- Optional, aber gerne gesehen sind Erfahrungen im Bereich Single-Sign-on-Infrastrukturen und mit der Software SimpleSAMLphp oder Keycloak.

Fühlen Sie sich auch zu einer Bewerbung eingeladen, wenn Sie noch nicht alle der genannten Anforderungen erfüllen.

Unser Angebot
- Flexible Arbeitszeiten und die Möglichkeit zu mobilem Arbeiten sowie 30 Tage Urlaub
- Ein modernes, vielfältiges und außergewöhnliches Arbeitsumfeld mit großer Nähe zu Wissenschaft und Forschung an der Schnittstelle mehrerer innovativer Technologiesektoren
- Eine interessante, vielseitige Tätigkeit in einem großen, international agierenden IT-Kompetenzzentrum
- Mitarbeit in einem kompetenten und engagierten Team
- Qualifizierung und Weiterentwicklung Ihrer Fähigkeiten
- Im öffentlichen Dienst übliche Sozialleistungen, wie z. B. Betriebsrente mit zusätzlicher Erwerbsminderungs- und Hinterbliebenenversorgung

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 **26.04.2021** über unser Online-Formular unter *https://s-lotus.gwdg.de/gwdgdb/ago/20210304.nsf/bewerbung*.

Fragen zur ausgeschriebenen Stelle beantwortet Ihnen:

Herr Sascha Krull
Tel.: 0551 201-2162
E-Mail: *sascha.krull@gwdg.de* oder

Herr Christof Pohl
Tel.: 0551 201-1878
E-Mail: *christof.pohl@gwdg.de*

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Personalia

ABSCHIED VON HANS-JÜRGEN GUTSCH

Herr Hans-Jürgen Gutsch hat nach fast 30-jähriger Tätigkeit die GWDG zum 28. Februar 2021 verlassen, um in den wohlverdienten Ruhestand zu wechseln. Er begann seine berufliche Laufbahn am 15. September 1980 beim Max-Planck-Institut für Experimentelle Medizin in Göttingen und wechselte dann am 1. Juli 1991 zur GWDG. Hier war er, zuletzt in der Arbeitsgruppe „IT-Infrastruktur" (AG I), für den technischen Betrieb der verschiedenen Rechenzentrumsstandorte zuständig. Zu seinen Kernaufgaben gehörten die Planung und Aufrechterhaltung des laufenden Betriebs der elektrischen Versorgung und Klimatisierung der Serverräume wie auch der Aufbau und die Ausstattung von Racks und die Koordination des Einbaus von Hardware. Herr Gutsch war ein leidenschaftlicher Ausbilder und bildete in den Berufsbildern Kommunikationselektroniker – Fachrichtung Informationstechnik und ab 2010 Elektroniker für Geräte und Systeme aus. Von 2001 bis 2019 war er Mitglied im Prüfungsausschuss bei der IHK Hannover. Als Sicherheitsbeauftragter der GWDG überwachte er die Einhaltung der einschlägigen Vorschriften der niedersächsischen Bauordnung und führte die Unterweisung für Arbeitssicherheit – und vertretungsweise auch für Brandschutz – durch. Herr Gutsch hat durch seine pragmatische und hilfsbereite Art so manchen komplizierten Hardwareeinbau „gerettet". Wir danken Herrn Gutsch für seine engagierte, langjährige Arbeit und wünschen ihm für seinen weiteren Lebensweg alles Gute.

Kasprzak

NEUE MITARBEITERIN GABRIELA SCHOPF

Seit dem 1. März 2021 hat Frau Gabriela Schopf Aufgaben in der Verwaltung der GWDG übernommen. Sie ist ausgebildete Bürokauffrau und unterstützt die Verwaltung insbesondere im Drittmittelbereich. Frau Schopf ist telefonisch unter 0551 201-26805 und per E-Mail unter *gabriela.schopf@gwdg.de* zu erreichen.

Suren

INFORMATIONEN:
support@gwdg.de
0551 201-1523

Mai bis Juli 2021

Academy

KURS	DOZENT*IN	TERMIN	ANMELDEN BIS	AE
QUICKSTARTING R: EINE ANWENDUNGSORIENTIERTE EINFÜHRUNG IN DAS STATISTIKPAKET R	Cordes	05.05. – 06.05.2021 9:00 – 12:00 und 13:00 – 15:30 Uhr	28.04.2021	8
WORKING WITH GRO.DATA	Király	11.05.2021 10:00 – 11:30 Uhr	10.05.2021	0
PHOTOSHOP FÜR FORTGESCHRITTENE	Töpfer	18.05. – 19.05.2021 9:30 – 16:00 Uhr	11.05.2021	8
INDESIGN GRUNDKURS – SCHWERPUNKT POSTERGESTALTUNG	Töpfer	02.06. – 03.06.2021 9:30 – 16:00 Uhr	26.05.2021	8
WORKING WITH GRO.DATA	Király	08.06.2021 10:00 – 11:30 Uhr	07.06.2021	0
ANGEWANDTE STATISTIK MIT SPSS FÜR NUTZER MIT VORKENNTNISSEN	Cordes	09.06. – 10.06.2021 9:00 – 12:00 und 13:00 – 15:30 Uhr	02.06.2021	8
OUTLOOK – E-MAIL UND GROUPWARE	Helmvoigt	24.06.2021 9:15 – 12:00 und 13:00 – 16:00 Uhr	17.06.2021	4
INDESIGN – AUFBAUKURS	Töpfer	29.06. – 30.06.2021 9:30 – 16:00 Uhr	22.06.2021	8
XUBUNTU-LINUX: XFCE-DESKTOP ALS ALTERNATIVE ZU POPULÄREN KOMMERZIELLEN BETRIEBSSYSTEMEN	Dr. Heuer	06.07.2021 9:00 – 12:00 und 13:30 – 15:30 Uhr	29.06.2021	4

KURS	DOZENT*IN	TERMIN	ANMELDEN BIS	AE
STATISTIK MIT R FÜR TEILNEHMER MIT VORKENNTNISSEN – VON DER ANALYSE ZUM BERICHT	Cordes	07.07. – 08.07.2021 9:00 – 12:00 und 13:00 – 15:30 Uhr	30.06.2021	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 im Kundenportal der GWDG (*https://www.gwdg.de*) einloggen. Wenn Sie zum Kreis der berechtigten Nutzer*innen der GWDG gehören und noch keinen GWDG-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
Aufgrund der aktuellen Corona-Situation finden zurzeit nahezu alle Kurse in einem geeigneten Online-Format und nicht als Präsenzkurse 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*. Falls bei einer ausreichend großen Gruppe Interesse besteht, könnten u. U. auch Kurse angeboten werden, die nicht im aktuellen Kursprogramm enthalten sind.

Kurz & knapp

Erreichbarkeit der GWDG am Maifeiertag, an Christi Himmelfahrt und um Pfingsten

Die Service-Hotline der GWDG ist am 01.05.2021, dem Maifeiertag, am 13.05.2021, Christi Himmelfahrt, sowie am 23.05. und 24.05.2021, den beiden Pfingstfeiertagen, telefonisch nicht erreichbar.

Falls Sie sich an diesen Tagen an die GWDG wenden möchten, erstellen Sie bitte eine Anfrage über unsere Support-Webseite unter *https://www.gwdg.de/support* oder schicken eine E-Mail an *support@gwdg.de*. Das dahinter befindliche Ticket-System wird auch an diesen Tagen von Mitarbeiter*innen der GWDG regelmäßig überprüft. Wir bitten alle Nutzer*innen, sich darauf einzustellen.

Das Rechenzentrum der GWDG bleibt für den Publikumsverkehr nach wie vor aufgrund der aktuellen Pandemiesituation bis auf Weiteres geschlossen.

Pohl

HPC-Statuskonferenz 2021 am 27.04.2021

Am 27. April 2021 veranstaltet das Bundesministerium für Bildung und Forschung zusammen mit der Gauß-Allianz und der Universität Göttingen sowie der GWDG eine Tagung zu aktuellen Entwicklungen im Bereich des High Performance Computings (HPC). Die Tagung wird als Online-Veranstaltung stattfinden.

Bei dieser Tagung werden aktuelle Entwicklungen in der HPC-Forschung und der HPC-Nutzung in Wissenschaft und Wirtschaft vorgestellt und diskutiert. Darüber hinaus ist auch die Kooperation mit europäischen Partnern in EuroHPC ein zentrales Thema. Im ersten Teil werden Ziele und Maßnahmen auf nationaler und europäischer Ebene vorgestellt sowie Highlights aus den Rechenzentren und aktuelle Anwendungen präsentiert. Im zweiten und dritten Teil stehen wissenschaftliche Ergebnisse und die Einbettung des Hoch- und Höchstleistungsrechnens in die Wissenschafts- und Innovationslandschaft im Vordergrund. Zum Abschluss ist eine Podiumsdiskussion vorgesehen, die Fragen der Teilnehmer*innen zu allen Themen der Tagung aufgreifen soll.

Weitere Informationen zur Veranstaltung sind unter dem URL *https://gauss-allianz.de/de/hpc-statuskonferenz-2021* zu finden.

Otto


GWDG
Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen