

30 Jahre Netzwerk – eine Evolutionsstory  
*30 Years of Network – an Evolutionary Story*

Erweiterte Cloudlösung für Anwender produktiv verfügbar  
*Extended Cloud Solution productively available for Users*

Nachhaltigkeitsaspekte bei der Entwicklung der Ginkgo-  
Software-Bibliothek

*Sustainability Aspects in Developing the Ginkgo Software Library*

## *Liebe Leserinnen und Leser,*

während die Künstliche Intelligenz und das Internet der Dinge zu den Trendtechnologien zählen und die digitale Transformation in den nächsten Jahren entscheidend prägen werden, ist das Cloud Computing in der Welt der IT-Zentren und Serviceprovider, aber auch bei den Endnutzern, voll angekommen. Kaum ein Anwender weiß mehr, wo seine digitalen Bilder, Termine und Video-Dateien wirklich physisch abgespeichert sind. Und auch die Sysadmins bauen keine stromfressenden Server mehr in ein Rack, wenn es darum geht einer Forschungsgruppe eine moderne Entwicklungsplattform bereitzustellen. Was früher eine mehrere Stunden dauernde Angelegenheit war, geschieht heute binnen Minuten und voll automatisiert. Für Hochschulen und Universitäten in Baden-Württemberg leistet dies seit Oktober 2018 der Dienst bwCloud SCOPE verteilt über vier Betriebsstandorte (S. 9).

Solche virtualisierten Lösungen wurden erst durch leistungsfähige Netzwerke und die Fortentwicklung des Internet möglich. Niemand würde z.B. einen Brief in einer Web-Anwendung schreiben, wenn die Buchstaben erst Sekunden nach dem Tastendruck auf dem Bildschirm folgen. Über die nun 30-jährige Entwicklung der Netzwerktechnik, in und außerhalb des KIT, und die Herausforderungen bei der Jagd nach immer schnelleren Transferraten schreibt auf Seite 4 unser langjähriger Abteilungsleiter Netze, Reinhard Strebler, der Ende 2018 in den wohlverdienten Ruhestand wechselte. Für sein großes Engagement in den letzten 30 Jahren bedanken wir uns sehr herzlich und wünschen ihm alles Gute.



So wie schnelle Netze für Cloud-Dienste, sind Supercomputer elementar für das wissenschaftliche Rechnen. Dafür optimierte und auf Hochleistungsrechnern lauffähige Softwaremodule sind neben smarten Web-Apps ebenfalls Themen des SCC. Im Bericht über das Gingko-Projekt (S. 24), erfahren Sie wie ein nachhaltiger Entwicklungsprozess Software mit hoher Qualität und Zuverlässigkeit hervorbringt.

Viel Freude beim Lesen.

*Martin Frank, Bernhard Neumair, Martin Nußbaumer, Achim Streit*

## *Dear reader,*

while artificial intelligence and the Internet of Things are among the trending technologies that will shape digital transformation in the years to come, the cloud has become today's mainstream technology in data centers and service providers as well as for consumers. Hardly anyone knows where her digital pictures, video files and other documents are actually stored. System administrators no longer install power-hungry servers when it comes to providing a research group with a modern software development platform. What used to take several hours happens today fully automated in a matter of minutes. For colleges and universities in Baden-Württemberg the bwCloud SCOPE service, which is distributed over four operating locations, is in operation since October 2018 (p. 9).

Such virtualized solutions are only possible with powerful networks and the development of the Internet. Nobody would for example write a document in a web application if the letters appear on screen seconds after typing. SCC's former network department head, Reinhard Strebler, who recently moved into his well-deserved retirement, writes about 30 years development of network technology and the challenge in the search for ever faster transfer rates in and outside KIT (p. 4). We would like to deeply thank Reinhard Strebler for his commitment and dedication over the last 30 years, and wish him all the best for the future.

Like fast networks for cloud services, supercomputers are fundamental for scientific computing. Optimized software modules that can be run on supercomputers as well as smart web applications are both research topics at SCC. In the report about Gingko (p. 24), the sustainable development process to develop high quality and reliable software is described.

Enjoy reading!

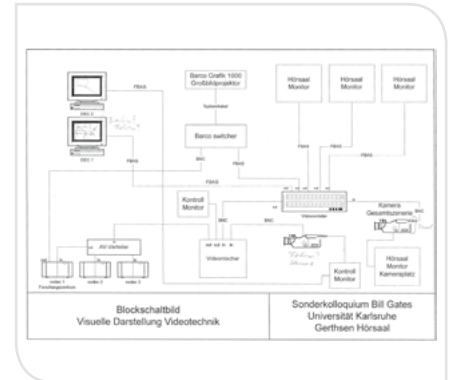
*Martin Frank, Bernhard Neumair, Martin Nußbaumer, Achim Streit*



# Inhaltsverzeichnis

## DIENTE UND INNOVATION

- 04 30 Jahre Netzwerk – eine Evolutionsstory
- 08 RDM@KIT: Professionelle Services für nachhaltiges Datenmanagement
- 09 Erweiterte Cloudlösung für Anwender landesweit produktiv verfügbar
- 13 Neue Speichersysteme für GridKa und die LSDF
- 16 Das neue SCC-Ticketsystem – erweiterte Möglichkeiten der Kommunikation
- 17 Smart studieren am KIT = mobil studieren
- 19 IT-Angriffe: GEMEINSAM das KIT schützen
- 20 Produktivsetzung des Anlageninventursystems im Universitätsbereich
- 21 Neuerungen von A-Z: Die Lernplattform -> einmal komplett umgekrempelt



## FORSCHUNG UND PROJEKTE

- 22 Der Helmholtz-Inkubator Information & Data Science
- 24 Nachhaltigkeitsaspekte bei der Entwicklung der Ginkgo-Software-Bibliothek
- 28 Effiziente Rechenlastverteilung für die Simulation von turbulenter Verbrennung auf Supercomputern
- 31 Helmholtz Analytics Framework
- 34 KIT-Zentrum Mathematics in Sciences, Engineering, and Economics (MathSEE) hat seine Arbeit aufgenommen
- 36 Aufbau einer AAI für die Helmholtz Data Federation
- 39 Rechenzeitanträge für den ForHLR am SCC
- 40 Das SCC prägt die European Open Science Cloud



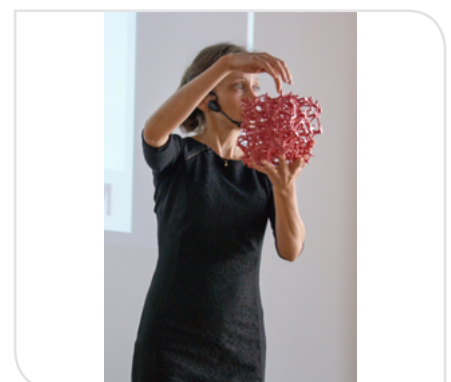
## STUDIUM UND WISSENSVERMITTLUNG

- 42 MINTernationaler Erfolg des KIT
- 43 MINT-EC Akademie CAMMP week im September in Karlsruhe
- 44 16. Internationale GridKa School – Computing and Science Fair



## VERSCHIEDENES

- 48 Erster Doctoral Science Slam am SCC
- 50 Reinhard Strebler geht in den Ruhestand
- 51 Neues aus den SCC-Abteilungen
- 51 Impressum & Kontakte



## 30 Jahre Netzwerk – eine Evolutionsstory

Mit Hilfe des Campusplans auf dem Smartphone zum Treffpunkt kommen, eine Präsentation mit einer internationalen Arbeitsgruppe online bearbeiten, Bilder und Publikationen versenden, Weiterbildungsangebote mittels Web-Seminaren nutzen und sich über Videokonferenzen in Projekten austauschen, das sind nur einige Beispiele, die unseren Alltag in einer digitalisierten und vernetzten Welt bestimmen. Hinter den smarten Apps und Anwendungen steckt heute eine Menge moderner Rechner- und Netzwerktechnik, ohne die ein erfolgreiches Forschen und Lehren in einer Wissenschaftseinrichtung nicht mehr möglich wäre. Dieser Artikel blickt über zwei SCC-News-Ausgaben auf spannende Entwicklungen in der Netzwerktechnik und deren Umsetzung am KIT und seinen Vorgängereinrichtungen zurück.

Reinhard Strebler

### Das Festnetz und die Technik

In den 1980er Jahren begann am KIT die Ära der standardkonformen Vernetzung von Computersystemen. Davor gab es bereits Remote-Verbindungen im Großrechnerbereich unter Einsatz proprietärer Protokolle. Mit der Verfügbarkeit des Industriestandards DIX, ein Vorläufer von Ethernet, entstanden erste lokale Datennetze in Instituten. Mit der Veröffentlichung des ersten Ethernet-Standards 10BASE5 [1] durch das Institute of Electrical and Electronics Engineers (IEEE) begann der Siegeszug dieser Protokolle. Es gab zwar noch StarLan (IEEE802.3, 1BASE5, 1 Mbit/s) [2] und Token Ring (IEEE 802.5, 4 Mbit/s) [3], jedoch war bald abzusehen, dass diese keine langfristige Weiterentwicklung erleben dürften. Bei Ethernet entsprechend 10BASE5 musste ein ca. 10mm dickes Koaxialkabel verlegt werden [Abb. 1]. Die maximale



Abb. 1: Yellow Cable

Segmentlänge betrug 500 Meter. In Abständen von mindestens 1,5 m konnten sogenannte Vampire Taps [Abb. 2] angebracht werden, an die Transceiver [Abb. 3] montiert werden konnten. Über AUI-Kabel (Attachment Unit Interface) wurden dann Endgeräte oder Netzwerkkomponenten angeschlossen. Auf Grund der rasanten Verbreitung von Ethernet war davon auszugehen, dass diese Technologie sich lange halten wird. Mit 10BASE2 gab es 1985 eine Erweiterung des Standards. Das nun einsetzbare, dünnere RG-58-Kabel war leichter zu verlegen und die Montage der Vampir Taps entfiel durch den Einsatz von T-Stücken. Allerdings waren die Segmentlängen auch deutlich reduziert (max. 185 m).



Abb. 2: Vampire Tap mit Mittelkontakt und Bohrwerkzeug



Abb. 3: 10BASE5 Transceiver

Ethernet war damals als „shared Medium“ definiert. Ein wesentliches Merkmal war CSMA/CD (Carrier Sense, Multiple Access/ Collision Detection) [3]. Dabei musste ein sendewilliger Teilnehmer zuerst hören, ob bereits Verkehr auf dem Medium „Kabel“ stattfand (Carrier Sense). Wenn das Medium frei war, durften alle sendewilligen Teilnehmer mit der Übertragung starten (Multiple Access). Kam es dabei zu einer Kollision, weil mehr als ein Teilnehmer zum selben Zeitpunkt Daten sendete, mussten alle Sender stoppen und die Übertragung nach einer zufälligen Wartezeit erneut starten. Der erreichbare Durchsatz war dadurch auf ca. 30 – 70% der verfügbaren 10 Mbit/s limitiert.



Im Standard IEEE 802.3 waren auch Repeater zur Kopplung der Segmente definiert. Der maximale Durchmesser eines Ethernet-Netztes wurde dadurch von 500 Meter auf 2.500 Meter erweitert. Auch beim Einsatz von Repeatern wurde der Traffic im gesamten Netz geflutet. Eine Lasttrennung war im Standard nicht vorgesehen.

Ab Mitte der 1980er Jahre gab es dann die ersten Ethernet-Bridges. Diese waren meistens mit 2 Anschlüssen ausgestattet. Hiermit war eine Lasttrennung zwischen Netzsegmenten möglich, indem geprüft wurde, ob die Übertragung eines Ethernet-Pakets erforderlich war. Waren sowohl Sender als auch Empfänger auf derselben Seite einer Ethernet-Bridge, wurde das Paket nicht weitergeleitet. Diese Bridges waren rein Software-basiert und damit nicht besonders performant.

Da beim Aufbau größerer Netze eine Lasttrennung mittels Ethernet-Bridges nicht zielführend ist und es bei Ethernet keine Standards für höhere Bandbreiten gab, musste für das Kernnetz eine andere Technologie gefunden werden. Hierfür wurde an den beiden großen Lokationen des KIT FDDI (Fiber Distributed Data Interface, ANSI (American National Standards Institute) Standard X3T9.5 [4]) eingesetzt. FDDI bot mit einem gegenläufigen Glasfaserdoppelring eine Bandbreite von maximal 2\*100 Mbit/s bei Nutzung beider Ringe. Fiel eine Verbindung oder ein Teilnehmer aus, wurde über einen optischen Bypass-Schalter der Ring wieder geschlossen. In diesem Modus stand aber nur die halbe Bandbreite zur Verfügung.

Im Jahr 1990 wurde mit 10BASE-T ein Standard für die Übertragung von 10Mbit/s-Ethernet über Twisted-Pair-Kabel etabliert. Diese mussten eine Qualität der sogenannten „Kategorie 3“ aufweisen. Hiermit wurde die Basis für eine synergetische Nutzung der letzten Meile sowohl für Ethernet als auch für Telefonie geschaffen. Umgehend entstand auch ein De-facto-Standard FOIRL (Fiber Optic Inter Repeater Link) zur Kopplung von Netzsegmenten über Glasfaser, der im Jahr 1993 zu dem IEEE-Standard 10BASE-F führte.

Zeitgleich mit der Verfügbarkeit der ersten standardisierten 100 Mbit/s-Ethernet-Technologie (1995) und der Definition von Voll-Duplex-Ethernet als Standard etablierte sich das sog. Switching im Markt. Das Voll-Duplex-Verfahren erlaubt das gleichzeitige Senden und Empfangen auf einem Twisted-Pair-Segment. Dadurch konnte nun 100% der physikalischen Bandbreite genutzt werden. Die ersten Ethernet-Switches hatten 8 bis 16 Ports und verfügten über spezielle Hardwarebeschleuniger (ASICs). Diese Hardwareteile waren insofern proprietär, weil jeder Switch-Hersteller eigene ASICs (Application-Specific Integrated Circuit) entwickelte. Bis heute gibt es einige wenige große Chip-Hersteller, deren Produkte sich in fast allen Netzwerkkomponenten finden.

Anfang der 1990er Jahren kam der FDDI-basierte Backbone an die Grenzen seiner Leitungsfähigkeit. Der große Hype im Netz-

werkbereich hieß damals ATM (Asynchronous Transfer Mode), eine Technologie, die Telefon- und Datennetze zusammenführen sollte. ATM basierte auf den hohen SONET/SDH-Bandbreiten (34 Mbit/s, 155 Mbit/s, 622 Mbit/s, 2.488 Gbit/s) und war gerade deshalb die große Hoffnung.

Auf Grund der erhofften Langlebigkeit der Technologie wurden zeitgleich sowohl am Forschungszentrum als auch an der Universität ATM-Netze im Backbone geplant und umgesetzt.

### Sorgenkind Westhochschule

Die „Außenstelle“ Westhochschule wurde anfangs über eine angemietete 64 Kbit/s Standleitung angebunden. Bald erfolgte die Hochtastung auf 2 Mbit/s. Versuche mit einer optischen Funkstrecke zur Anbindung der Westhochschule mit 10 Mbit/s waren zwar nicht völlig erfolglos, aber atmosphärische Einflüsse wie Regen und Nebel führten regelmäßig zu Ausfällen dieser Strecke. Temperaturschwankungen hatten ebenfalls Auswirkungen auf die Verfügbarkeit und erforderten regelmäßig Nachjustierungen der Sendelaser. Nach langer Planung wurde 1995 ein Glasfaserkabel in Betrieb genommen, das die Westhochschule mit dem Campus Süd des KIT verband. Damit war die Möglichkeit gegeben, diese „Außenstelle“ mit entsprechender Bandbreite an das Kernnetz anzubinden.

### Bill Gates an der Uni – we can do IT!

Im Jahr 1994 förderte Microsoft das University Support Centre [6] (heute Education Support Centre Deutschland [7]), am Institut für Programmstrukturen und Datenorganisation (Prof. Tichy) eine gegründete Initiative zur Unterstützung von Studierenden und Hochschulen in Deutschland. Auf Grund des überregionalen Erfolgs dieser Initiative fragte Microsoft bei der Universität Karlsruhe (TH) 1996 an, ob Bill Gates vor den Studierenden die Firmenstrategie von Microsoft darstellen und erläutern könne. Am 8. Februar 1996 fand dann das große Ereignis statt, das nach langer und intensiver Vorbereitung reibungslos klappen sollte. Damit möglichst viele Studierende Bill Gates hören und sehen konnten, wurde mittels ATM ein Videokonferenzprojekt mit den damals begrenzten Mitteln aufgesetzt. Bill Gates präsentierte im Gerthsen-Hörsaal, der zu jenem Zeitpunkt die meisten Sitzplätze bot. Die Präsentation wurde per ATM in weitere Hörsäle an der Universität Karlsruhe (TH) übertragen (Gaede, NTI), sowie zum großen Hörsaal am FTU<sup>1</sup> des Forschungszentrums und zur TU Berlin. An den so zugeschalteten Orten konnten die Studierenden am Ende der Veranstaltung auch direkte Rückfragen an Bill Gates richten.

Eingesetzt wurden ATM-Switches von Fore Systems (ASX200BX) und Cisco (Lightstream 1010). Hierüber wurden Point-to-Multi-

<sup>1</sup> Fortbildungszentrum für Technik und Umwelt

point PVCs (permanent virtual circuit) konfiguriert, die vom Gerthsen-Hörsaal ausgingen. Für die Rückfragekanäle wurden Point-to-Point VPCs (Virtual Port Channel) konfiguriert [Foto Seite 50]. Die für die Digitalisierung der Video- und Audioströme erforderlichen Codecs wurden mittels Sun und DEC Alpha Workstations realisiert. Der im Gerthsen-Hörsaal aufgezeichnete Audio- und Videostrom wurde gleichzeitig an alle Außenstellen übertragen. Bei den Rückfragen wurde dann ausgewählt, welche Außenstellenmikrofone auf das Audiosystem im Gerthsen-Hörsaal aufgeschaltet werden sollten [Abb. 6]. Für die ATM-Verbindung zur TU Berlin wurde der Telekom-Dienst B-ISDN (Breitband ISDN) genutzt und eine Bandbreite von 34 Mbit/s bereitgestellt. Mit diesem Dienst wollte die Telekom einen Bandbreitenvermittlungsdienst etablieren. Verbindungswünsche mussten mit einer Vorlaufzeit von einigen Arbeitstagen per Fax gestellt werden. Trotz dieser schwierigen Bedingungen klappte alles reibungslos.

### Das Virtuelle Rechenzentrum

Im Jahr 1996 wurde das „Virtuelle Rechenzentrum“ als Kooperation zwischen FZK und Uni für das Vektor- und Parallelrechnen gegründet. Beide Rechenzentren waren durch angemietete Glasfasern mit 155 Mbit/s-ATM verbunden. 80 Mbit/s wurden für Datendienste, der Rest als Reserve für spontane Anwendungen wie Videoübertragungen genutzt.

Für die Nutzung von ATM im Netzwerkbereich gab es zwei konträre Ansätze: LAN-Emulation und Classical IP. LAN-Emulation bildete die Eigenschaften eines Ethernet-Netzwerks auf dem ATM-Layer ab. Hierzu waren Steuerungsinstanzen erforderlich, da Ethernet verbindungslos, ATM aber verbindungsorientiert ist. Mit LAN-Emulation konnten parallele, logische LAN-Strukturen implementiert werden, was zu diesem Zeitpunkt mit Ethernet-on-the-wire nicht standardkonform abgebildet werden konnte. Im Jahr 1996 definierte die IEEE mit dem Standard 802.1Q sowohl die Basis für eine Virtualisierung im LAN (Schicht 2 des OSI-Modells) als auch für die Einführung einer Priorisierung auf Netzwerkschicht 2. Die IETF (Internet Engineering Task Force, Standardisierungsgremium der TCP/IP-Protokolle) definierte im RFC 1577 mit Classical IP eine Technik, mit der IP-Pakete direkt über ATM verschickt werden konnten, ohne dass diese vorher in ein Ethernet-Paket verpackt werden mussten.

In beiden Varianten mussten die relativ großen Ethernet- bzw. IP-Pakete zerhackt werden, damit die Daten in ATM-Zellen verschickt werden konnten. Die ATM-Zellen enthielten einen 5-Byte Header und boten 48 Byte für die Nutzinformation. Auf Grund dieser fixen Zell-Struktur erfolgte das Weiterleiten in den Vermittlungsstationen (ATM-Switches) höchst performant. Gleichzeitig war diese Technik aber der Todesstoß für die Nutzung von ATM in Hochgeschwindigkeitsnetzen. Die Entwicklung und Produktion von SAR-Bausteinen (Segmentation and Reassembly) für Geschwindigkeiten über 622 Mbit/s war derart aufwändig und teuer, dass eine weitere Verbreitung von ATM in Hochgeschwindigkeitsnetzen unwirtschaftlich war. Eine kostengünstige Alternative bot die Standardi-

sierung von 1 Gbit/s-Ethernet zum Jahrtausendwechsel. Wegen des raschen Preisverfalles wurde der weitere Ausbau des Netzes mit ATM gestoppt und eine Migration zu einem Gigabit-Ethernet-Netz eingeleitet. Bei den dort eingesetzten optischen Schnittstellen setzte eine erfreuliche Entwicklung ein. In der Vergangenheit waren diese generell Bestandteil der Interfacemodule der Netzkomponenten. Durch den Modultyp war festgelegt, für welche Art von Glasfaserverbindungen (Multi Mode oder Single Mode) es benutzt werden konnte. Bei optischen 1 Gbit/s-Schnittstellen wurden standardisierte Umsetzer, sog. GBICs (Gigabit Interface Converter), eingesetzt. Die Gigabit Ethernet Alliance spezifizierte die elektrische Schnittstelle zwischen optischem Modul und dem Einschubslot der Netzwerkkomponente. Um die Packungsdichte der 1 Gbit/s-Schnittstellen der Netzkomponenten zu erhöhen, wurden kleinere Module (SFP, Small Form Factor Pluggable) spezifiziert, die auch heute noch Stand der Technik sind [Abb. 4].



Abb. 4: GBIC und SFP

Mit dem Siegeszug der Ethernet-Technik setzte eine rasante Entwicklung ein. Im Jahr 2002 wurde der IEEE-Standard für optische Übertragungen gemäß 10GBASE-x fertiggestellt, vier Jahre später die elektrische Übertragung über Twisted Pair genormt. Allerdings waren damals die Optiken für 10GBASE-SR (für Multimode Glasfasern) und 10GBASE-LR (für Singlemode-Glasfasern) mit Kosten bis ca. 10.000€ recht teuer. In den folgenden Jahren wurden die optischen Module verkleinert (XENPAK, X2, SFP+) [Abb. 5], was auch zu einem geringeren Stromverbrauch und damit einer geringeren Wärmebildung führte. Damit konnte die Anzahl von 10 Gbit/s-Schnittstellen je Modul bzw. Einschubkarte deutlich erhöht werden. Der folgende Preisverfall beschleunigte die Verbreitung von 10 Gbit/s massiv.



Abb. 5: XENPAK, X2 und SFP+

## Auf dem Weg zum KITnet

In der nächsten Ausgabe der SCC-News stehen die Herausforderungen im Fokus, die der Zusammenschluss von Forschungszentrum und Universität Karlsruhe an ein gemeinsames Netz stellen. Das KITnet ist dabei die Grundlage für das leistungsfähige und zuverlässige wissenschaftliche Rechen- und Datenzentrum des KIT, das Steinbuch Centre for Computing. Lesen Sie dann über den Aufbau einer campusübergreifenden Funknetzinfrastruktur, die Sprach- und Datenintegration, den Ausbau von Projektnetzen, Ergebnisse aus Forschungs- und Innovationsprojekten sowie zu Management-Tools und Serverdiensten für das KITnet.

- [1] [de.wikipedia.org/wiki/Ethernet](http://de.wikipedia.org/wiki/Ethernet)
- [2] [en.wikipedia.org/wiki/StarLAN](http://en.wikipedia.org/wiki/StarLAN)
- [3] [de.wikipedia.org/wiki/Carrier\\_Sense\\_Multiple\\_Access/Collision\\_Detection](http://de.wikipedia.org/wiki/Carrier_Sense_Multiple_Access/Collision_Detection)
- [4] [informat.com/articles/article.aspx?p=131227&seqNum=5](http://informat.com/articles/article.aspx?p=131227&seqNum=5)
- [5] [de.wikipedia.org/wiki/FCAPS](http://de.wikipedia.org/wiki/FCAPS)
- [6] [ps.ipd.kit.edu/415\\_2931.php](http://ps.ipd.kit.edu/415_2931.php)
- [7] [www.escde.net/](http://www.escde.net/)

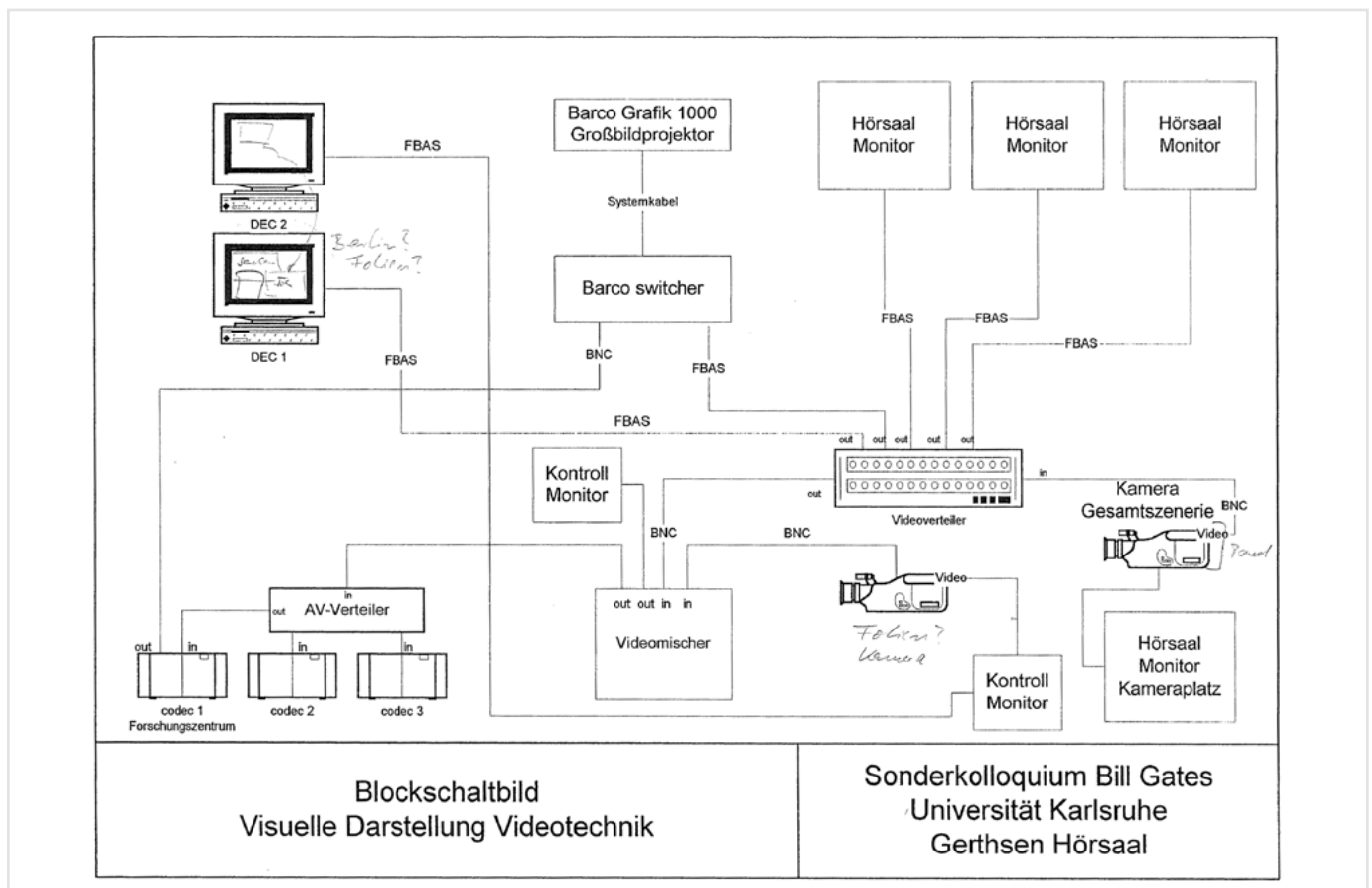


Abb. 6: Bill Gates Video Technik

### 30 Years of Network – an Evolutionary Story

Using the campus map on the smartphone to find your meeting point, editing a presentation on-line with an international working group, sharing publications and other scientific data, education offerings by means of web seminars and participating in projects in a video conference, are just a few examples which are inseparable from our everyday life in a digital and networked world.

Hidden behind these applications is a lot of modern network and computer technology, which enables the outstanding research and teaching at KIT. In this article Reinhard Strebler, outgoing and long-term head of the network department of SCC, recalls exciting developments in networking technology and their implementation over the last decades.



## RDM@KIT: Professionelle Services für nachhaltiges Datenmanagement

Nachhaltiges Forschungsdatenmanagement (FDM) gewinnt an Bedeutung in allen Wissenschaftsbereichen. Ein strategisches Ziel des KIT ist es, Infrastrukturen und Dienste aufzubauen, mit denen Forschungsdaten gemäß guter wissenschaftlicher Praxis erhoben, archiviert und öffentlich zugänglich und somit nachnutzbar gemacht werden können. An diesem Ziel, das in der Forschungsdatenrichtlinie des KIT festgehalten ist, wird systematisch gearbeitet. Das KIT hat sich frühzeitig mit Forschungsdatenmanagement befasst und sich dadurch unter deutschen Universitäten einen Wissensvorsprung und eine Vorreiterrolle erarbeitet.

*Felix Bach*



Seit 2014 unterstützt das Expertenteam RDM@KIT Forschende des KIT beim nachhaltigen Umgang mit Forschungsdaten [1]. Das Team, in dem das SCC und die KIT-Bibliothek federführend sind, besteht auch aus Mitgliedern des KIT-Archivs, der Forschungsförderung und des Innovations- und Relationsmanagement.

RDM@KIT bietet einen 360-Grad-Support von der Beratung bis zur operativen Unterstützung für sämtliche Phasen des Research Data Life Cycle (siehe Bild), angefangen von der Antragstellung wissenschaftlicher Projekte bis hin zur

Publikation, Archivierung und Nachnutzung von Forschungsdaten. Forschende können hier erfahren, welche Dienste und Infrastrukturen intern und extern zur Verfügung stehen. Zugleich werden Tools und Infrastrukturen entwickelt und ausgebaut, um die bislang vorhandenen Lücken im FDM-Prozess zu schließen, die in den bwFDM-Projekten und im Projekt bwDIM identifiziert wurden. Beispielsweise bietet der Dienst bwDataArchive [2] Wissenschaftlerinnen und Wissenschaftlern eine Langzeitarchivierung ihrer Daten. Der Dienst re3data [3] ist hingegen eine Plattform, um gezielt geeignete Forschungsre-

positorien zu recherchieren und so einen Überblick über Best-Practice-Beispiele zu erhalten.

Die Team-Mitglieder engagieren sich darüber hinaus in verschiedenen Arbeitsgruppen, z.B. in der Research Data Alliance (RDA), und beteiligen sich an vielen nationalen und internationalen Initiativen, die sich mit dem Thema Forschungsdatenmanagement befassen. Innerdeutsch profiliert sich insbesondere das Land Baden-Württemberg als Innovationsmotor im Zusammenhang mit eScience. An vielen der geförderten Projekte ist das KIT beteiligt und hat die Federführung für die Entwicklung und das Betreiben des Informationsportals Forschungsdaten.info [4] übernommen. Es soll Universitäten, Hochschulen und Forschungseinrichtungen bundesweit als Informationsplattform für den Umgang mit Forschungsdaten dienen.

**Kontakt:** [contact@rdm.kit.edu](mailto:contact@rdm.kit.edu)

### Web-Links:

- [1] [www.rdm.kit.edu](http://www.rdm.kit.edu)
- [2] [www.rda.kit.edu](http://www.rda.kit.edu)
- [3] [www.re3data.org](http://www.re3data.org)
- [4] [www.Forschungsdaten.info](http://www.Forschungsdaten.info)

### RDM@KIT: Professional Services for Sustainable Data Management

Sustainable research data management (RDM) is gaining importance in all scientific fields. A strategic goal of KIT is the establishment of infrastructures and services that allow research data to be recorded, archived, and made publicly accessible and thus re-usable according to good scientific practice. This goal, which is laid down in the KIT Research Data Policy, is being systematically pursued. KIT dealt with research data management at an early stage, thereby gaining a knowledge lead and a pioneering role among German universities.

# Erweiterte Cloudlösung für Anwender landesweit produktiv verfügbar

bwCloud SCOPE ist der föderierte Cloud-Service für die Universitäten und Hochschulen im Land Baden-Württemberg. Um eigene Projekte oder Arbeitsgruppen mit virtualisierter IT-Infrastruktur auszustatten, können Nutzer eigene virtuelle Maschinen starten und betreiben. Der Dienst steht Wissenschaftlern, Studenten und technischen Angestellten zur Verfügung. Der Nutzungsumfang ist über zwei verschiedene Berechtigungsstufen geregelt. Diese müssen die IT-Beauftragten der Organisationseinheiten in der Gruppenverwaltung vergeben, bevor die Registrierung möglich ist. So stellt bwCloud SCOPE<sup>1</sup> als Landesdienst eine Virtualisierungsumgebung auf OpenStack-Basis bereit.

Matthias Knoll, Klaus Scheibenberger, Janne Chr. Schulz<sup>2</sup>

## 1. Einführung



Das Projekt bwCloud - Standortübergreifende Servervirtualisierung, das Ende 2016 abgeschlossen wurde, hatte zum Ziel, ein Konzept zur föderierten Virtualisierung von Servern und Diensten zu entwickeln.

Das Nachfolgeprojekt bwCloud SCOPE, mit einer Laufzeit bis Ende 2019, hat sich nahtlos daran angeschlossen. Mit diesem Projekt soll eine föderierte Virtu-

alisierungsumgebung als Landesdienst bwCloud SCOPE etabliert werden.

Die Fokussierung auf drei Nutzergruppen spiegelt das Akronym SCOPE wider: Es benennt die drei adressierten Bereiche **Science**, **Operations** und **Education**, d.h. der neue Landesdienst unterstützt:

- (SC) Anforderungen aus der Forschung und Wissenschaft, beispielsweise für projektspezifische Systeme oder als Umgebungen für umfangreichere Analyse- oder Berechnungsarbeiten

- (OP) Klassische, innerbetriebliche Anforderungen (z.B. Webserver) von Infrastruktureinrichtungen wie Rechenzentren an den Hochschulen und Universitäten
- (E) Anforderungen aus der Lehre zur Umsetzung von Lehr- und Lernszenarien, beispielsweise für Praktika, Übungen oder Abschlussarbeiten

## 2. bwCloud SCOPE Infrastruktur

Ende 2017 wurde mit der Ausschreibung der notwendigen System- und Speicherressourcen der Aufbau einer neuen Cloud-Infrastruktur initiiert. Diese ist, verglichen mit dem ersten Prototyp, wesentlich leistungsfähiger - mehr als 100 Rechen-Knoten mit Zugriff auf über 2,5 Petabyte Speicher (brutto) bieten Kapazität für tausende virtueller Maschinen (Tabelle 1).

Der Aufbau dieser Infrastruktur war Anfang Q2/2018 abgeschlossen. Nach einem Test- und Pilotbetrieb startete die Migration von Projekten in die neue Infrastruktur. Anfang Oktober 2018 konnte auch diese Phase erfolgreich abgeschlossen (Abbildung 2) und die System- und Speicherressourcen für die Altumgebung außer Betrieb genommen werden. Seit Mitte Oktober 2018 ist bwCloud SCOPE als Nachfolgedienst für bestehende und neue Nutzer vollständig einsatzbereit.



Abbildung 1: Die vier Betriebsstandorte (gelb) des föderierten Cloud-Dienstes bwCloud SCOPE

<sup>1</sup> bw-cloud.org

<sup>2</sup> Universität Mannheim

bwCloud Standort	Anzahl Rechnerknoten	Speicherplatz	CPUs / Kerne
Karlsruhe (KIT/SCC)	27	504 TByte*	27 / 432
Ulm (kiz)	36	1.200 TByte*	36 / 576
Freiburg (Rechenzentrum)	27	504 TByte*	27 / 432
Mannheim (Rechenzentrum)	27	504 TByte*	27 / 432
Gesamt für Produktivbetrieb	117	2.712 TByte*	117 / 1872
* = Brutto			

Tabelle 1: Die aktuelle Ausstattung an den vier Betriebsstandorten

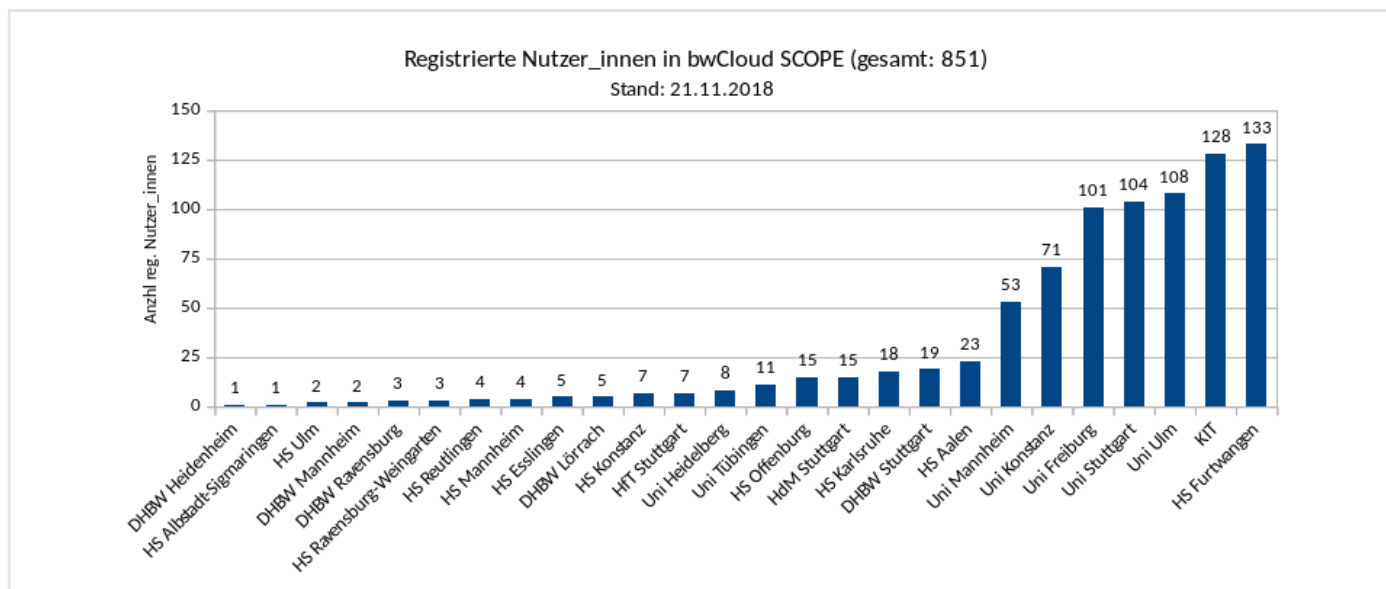


Abbildung 2: Nutzungsprofil zum Ende der Migrationsphase

### 3. Registrierung für den Dienst

#### A. Berechtigungssteuerung via Entitlements



Der Landesdienst bwCloud SCOPE ist in das föderierte Identitätsmanagement der baden-württembergischen

Hochschulen bwIDM<sup>3</sup> eingebunden. Damit Nutzer die Cloud-Infrastruktur verwenden können, ist eine Registrierung erforderlich. Dafür müssen die am bwIDM-Verband teilnehmenden Einrichtungen ihren Nutzern in den lokalen Identitätsprovidern sogenannte Entitlements zuordnen. Um sie für unterschiedliche Leistungsstufen des Dienstes freischalten zu können, wurden zwei verschiedene

Entitlements definiert: bwCloud-Basic und bwCloud-Extended.

Je nach individuell zugeordnetem Entitlement ist die verfügbare Kapazität (Quota) für das jeweilige bwCloud SCOPE-Dienstkonto unterschiedlich und damit auch die Nutzungsmöglichkeit (s. Abschnitt 5. Nutzung von bwCloud SCOPE).

Das Entitlement bwCloud-Basic umfasst Kapazitäten zum Betrieb einer einfachen virtuellen Maschine. Es bietet einen unkomplizierten Zugang zur bwCloud, damit sich neue Nutzer mit der Umgebung vertraut machen können.

Das Entitlement bwCloud-Extended richtet sich dagegen an die Power-User und berechtigt zur Nutzung von deutlich mehr Ressourcen. Es richtet sich vor allem an diejenigen, die bereits Erfahrung mit Cloud-

Umgebungen haben. Also Nutzer, die wissen, welche Ressourcen sie brauchen und wie sie diese am besten einsetzen.

Bei der Registrierung für den Dienst werden die Entitlements geprüft. Ist eines der beiden Entitlements vorhanden, wird die Registrierung freigeschaltet und ein Account in bwCloud SCOPE automatisch angelegt (s. Abbildung 3).

#### B. Berechtigungsvergabe am KIT

Am KIT werden die Entitlements über dessen Identitätsmanagementsystem (KIT-IDM) bzw. die damit verbundene Gruppenverwaltung zugeordnet. Angelehnt an den seit einigen Jahren produktiven Landesdienst bwSync&Share ermöglicht auch hier eine entsprechend angelegte Gruppenstruktur (s. Abbildung 3a) die dezentrale Verwaltung der Entitlements. Diese

<sup>3</sup> Siehe auch [www.bwidm.de](http://www.bwidm.de)





Abbildung 3: Entitlements werden dem Nutzer lokal zugeordnet

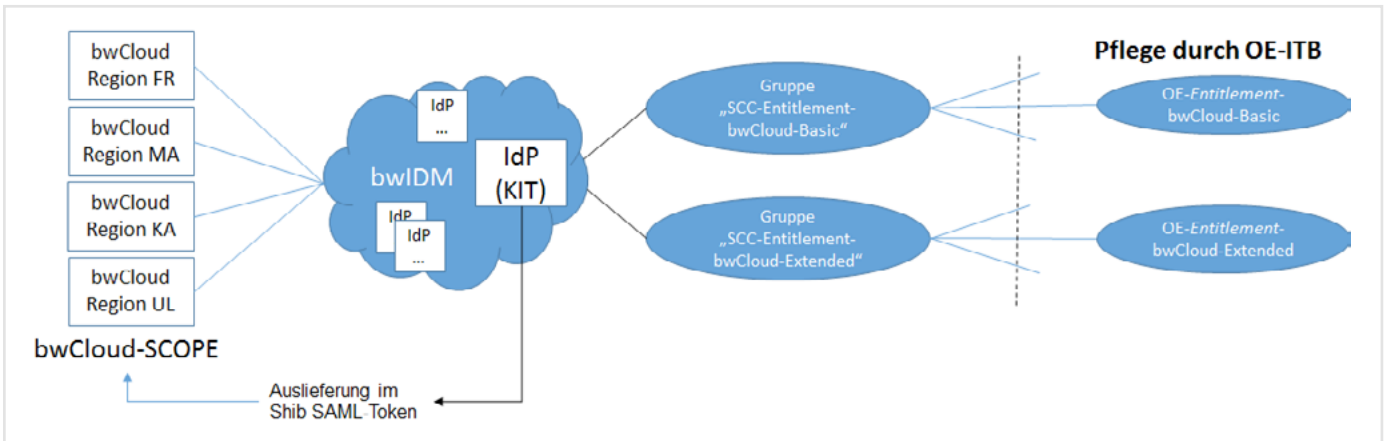


Abbildung 3a: Verwaltung der Entitlements in der zentralen Gruppenverwaltung mit automatischer Vergabe der SCC-Entitlements.

Struktur ist mit dem Identity Provider (IDP) in bwIDM gekoppelt, d.h. die Vergabe der Entitlements erfolgt in den Organisations-einheiten (OE) des KIT. Jede OE kann so in ihrem Kontext entscheiden, wem welches Entitlement zugeordnet wird.

#### 4. Nutzung der bwCloud SCOPE

##### A. Zusammenhang von Flavors und Entitlements

Zur besseren und einfacheren Verwaltung der bwCloud SCOPE-Ressourcen wurden sogenannte Flavors (Varianten) definiert (s. Tabelle 2). Bei bwCloud

SCOPE versteht man unter einem Flavor die Menge der nutzbaren virtuellen CPUs und die Größe des Hauptspeichers (RAM). Insgesamt kann ein Nutzer nur so viele Instanzen virtueller Systeme starten, dass die Summe der verwendeten Flavors die Gesamt-Quota nicht übersteigt. Nutzer können so individuelle Zusammenstellungen von virtuellen Maschinen starten.

##### Das Entitlement bwCloud-Basic

Mit diesem Entitlement erhält der Nutzer die Quota: 1 vCPU, 1.024 Megabyte RAM, 50 Gigabyte Festplattenspeicher

sowie eine IPv4- und IPv6-Adresse für die Netzwerkanbindung. Damit kann er entweder eine Instanz des Flavors Nano oder Tiny aufsetzen. Die Laufzeit dieser virtuellen Systeme ist auf sechs Monate begrenzt. Etwa vier bis sechs Wochen vor Ablauf der Laufzeit wird der Nutzer über die Löschung der Systeme per E-Mail informiert und kann so für die Sicherung seiner Daten sorgen.

##### Das Entitlement bwCloud-Extended

Mit dem Entitlement bwCloud-Extended sind 8 Instanzen, 16 vCPUs, 16 Gigabyte

	Flavor	vCPUs / RAM
Kostenfrei Entitlement "bwCloud-Basic"	Nano	1 / 0,5 Gigabyte
	Tiny	1 / 1 Gigabyte
Kostenpflichtig Entitlement "bwCloud-Extended"	Small1	1 / 2 Gigabyte
	Small2	2 / 2 Gigabyte
	Medium	2 / 4 Gigabyte
	Large	4 / 8 Gigabyte
	Xlarge	8 / 16 Gigabyte
	XXLarge	16 / 32 GigaByte

Tabelle 2: Ressourcen-Zusammenstellung über sog. Flavors

RAM, 128 Gigabyte Festplattenspeicher und zwei IPv4/IPv6-Adressen verbunden. Damit ist es möglich, virtuelle Maschinen auch mit größeren Flavors zu starten und diese ohne zeitliche Limitierung zu betreiben. Sollten die Quota-Einstellungen nicht ausreichen, kann jederzeit eine Erhöhung bei der bwCloud-Betriebsgruppe angefragt werden (s. [bw-cloud.org/q/t](http://bw-cloud.org/q/t); s.u. Abschnitt Ticketsystem).

### Kosten- und Leistungsverrechnung

Wie in der Tabelle 2 angedeutet, ist geplant, für die zukünftige Nutzung des Landesdienstes bwCloud SCOPE ein Kostenmodell einzuführen. Dieses sieht derzeit vor, dass die Nutzung mit dem Entitlement bwCloud-Basic kostenfrei bleibt und dass für die Nutzer mit Entitlement bwCloud-Extended eine Kosten- und Leistungsverrechnung erstellt wird. So unterstützt die Einführung der beiden Entitlements die Heimatstandorte und die bwCloud-Betriebsgruppe bei diesem nächsten großen Schritt.

Denn die Einführung einer landesweiten Kosten- und Leistungsverrechnung bedarf einiger Abstimmung und Rücksprachen. Diese Prozesse sind langwierig, da viele Stellen involviert sind und viele Fragen geklärt werden müssen. In der aktuellen Phase von bwCloud SCOPE sind daher

sowohl die Kostenstruktur als auch deren Abrechnung noch nicht etabliert. Demzufolge können die Nutzer, z.B. auch am KIT, die Dienstvarianten bwCloud-Basic und bwCloud-Extended kostenfrei nutzen.

### B: Betriebssysteme

Die bwCloud-Betriebsgruppe stellt aktuelle und gepflegte Betriebssystem-Images bereit. Diese Images lassen sich per Mausklick in der nutzer-eigenen Web-Administrationsoberfläche<sup>4</sup> starten. Dort stehen derzeit jedem Nutzer fertige Systemabbilder mit CentOS 7, Debian 9, Fedora Cloud Base 28, OpenSUSE Leap 15, Ubuntu 18.04 und Ubuntu 18.04 Minimal zur Verfügung<sup>5</sup>.

Technisch lassen sich alle aktuellen Windows-Varianten in der bwCloud betreiben, wenn der Nutzer für eine passende Lizenz sorgt. Derzeit müssen sie dazu jedoch ein eigenes, manuell vorbereitetes Windows-Image bereitstellen und in das System einfügen.

Auch andere Betriebssysteme mit der Architektur x86-64 sind so nutzbar. Dies erfordert, dass der Nutzer ein passendes Image manuell vorbereitet<sup>6</sup> und dann über die Nutzeroberfläche in die Cloud lädt sowie startet.

### C: Festplattenspeicher

Zur Speicherung von Daten können den virtuellen Instanzen sogenannte Volumens (Festplattenlaufwerke) zugeordnet werden. Diese sind auf die in der Quota definierte Speicherkapazität begrenzt und persistent.

**Weitere technische Details finden Nutzer und Interessierte auf der Projekthomepage unter [www.bw-cloud.org](http://www.bw-cloud.org).**

### 5. Ticketsystem

Als Support für die Nutzer steht das Ticketsystem des bw-Support-Portals zur Verfügung. Über dieses Portal können für alle bw-Dienste Probleme und Anfragen gemeldet werden. Das Team der bw-Cloud bearbeitet darüber alle Anfragen der Nutzer. Dazu gehören auch die oben erläuterten Quota-Anfragen für größere Projekte und Teams.

<sup>4</sup> [portal.bw-cloud.org](http://portal.bw-cloud.org)

<sup>5</sup> Alle Infos siehe [www.bw-cloud.org/de/bwcloud\\_scope/images](http://www.bw-cloud.org/de/bwcloud_scope/images)

<sup>6</sup> siehe [docs.openstack.org/image-guide/create-images-manually.html](http://docs.openstack.org/image-guide/create-images-manually.html)



### Extended Cloud Solution for Users in Baden-Württemberg available

bwCloud SCOPE is the federated cloud service for universities and polytechnics in Baden-Württemberg. Users can start and run their own virtual machines to equip their own projects or work groups with virtualized IT infrastructure. The service is available to scientists, students and technical employees. The scope of use is regulated by two different authorization levels. These must be assigned by the IT representatives of the organizational unit in the central group administration portal before registration is possible. Thus, bwCloud SCOPE provides a virtualization environment based on OpenStack as a federal state service.

## Neue Speichersysteme für GridKa und die LSDF

Die Abteilung Scientific Data Management des SCC betreibt derzeit zwei Multi-Petabyte-Speichersysteme mit einer Gesamtkapazität von 44 Petabyte zur Speicherung und performanten Verfügbarmachung wissenschaftlicher Daten. Diese befinden sich in der Large Scale Data Facility sowie im Grid Computing Centre Karlsruhe (GridKa) und stehen Nutzern verschiedener Wissenschafts-Communities in Baden-Württemberg und der Helmholtz-Gemeinschaft sowie den an GridKa beteiligten Teilchen und Astroteilchenphysik-Experimenten zur Verfügung.

Jan Erik Sundermann

Die Abteilung Scientific Data Management (SDM) des SCC betreibt derzeit zwei Multi-Petabyte-Infrastrukturen zur Verarbeitung und Speicherung großer Forschungsdaten. Dies sind die Large Scale Data Facility (LSDF) [1] und das Speichersystem des Grid Computing Centre Karlsruhe (GridKa) [2]. Die LSDF bietet Nutzern des KIT und deren Projektpartnern in Baden-Württemberg Zugriff auf einen Datenspeicher, der insbesondere für die Speicherung von wissenschaftlichen Messdaten und Simulationsergebnissen datenintensiver Wissenschaftsdisziplinen vorgesehen ist. Zu den Hauptnutzern der LSDF gehören neben anderen die Institute für Meteorologie und Klimaforschung, Photonenforschung und Synchrotronstrahlung sowie Toxikologie und Genetik des KIT. GridKa ist ein Daten- und Computing-Center für Teilchen- und Astroteilchenphysik-Experimente. Es stellt als deutsches Tier1-Zentrum im Rahmen des Worldwide LHC Computing Grid (WLCG) Ressourcen für alle vier LHC-Experimente bereit, speichert darüber hinaus Daten für das Pierre Auger Observatory und wird ab 2021 als Raw-Data-Center für das BELLE-II-Experiment fungieren.

Im Verlauf des Jahres 2016 wurden in der LSDF und im GridKa neue Speichersysteme beschafft und in der ersten Jahreshälfte 2017 in Betrieb genommen. Diese ersetzen und erweitern unterschiedliche alte Speichersysteme und vereinen in der LSDF die darauf betriebenen Dienste LSDF-DIS und bwFileStorage im neuen Dienst LSDF Online Storage [11]. Bei beiden Systemen handelt es sich um GxFS Storage Appliances der Firma NEC [4]. Das Design beider Systeme erlaubt es, diese flexibel sowohl in ihrer Größe als auch in ihrer Leistung zu skalieren. Als Software-defined Storage Layer kommt IBM Spectrum Scale [5] zum Einsatz. Der Speicher ist in wenige, sehr große Dateisysteme unterteilt, die es einerseits erlauben, den Speicher einfach zu verwalten, und andererseits ermöglichen, ihn für unterschiedliche Anwendungsszenarien, seien es unterschiedliche Experimente oder Funktionen wie Tape-Buffer, zu optimieren. Die Speichersysteme haben aktuell eine nutzbare Kapazität von 34 Petabyte in GridKa und 10 Petabyte in der LSDF. In Benchmarks wurden die kombinierten maximalen Lese-Schreib-Geschwindigkeiten mit 100 Gigabyte/s bzw. 25 Gigabyte/s gemessen.

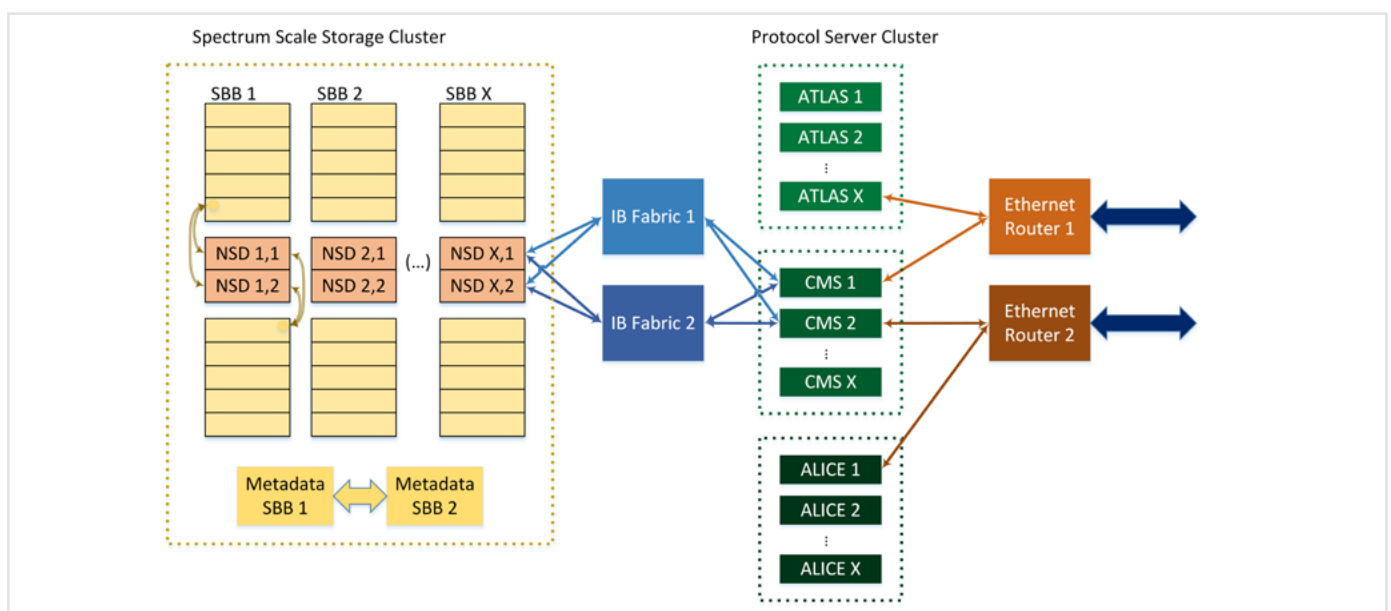


Abb. 1: Schematischer Aufbau des GridKa-Speichersystems, bestehend aus einem Storage-Cluster und mehreren Protokoll-Clustern



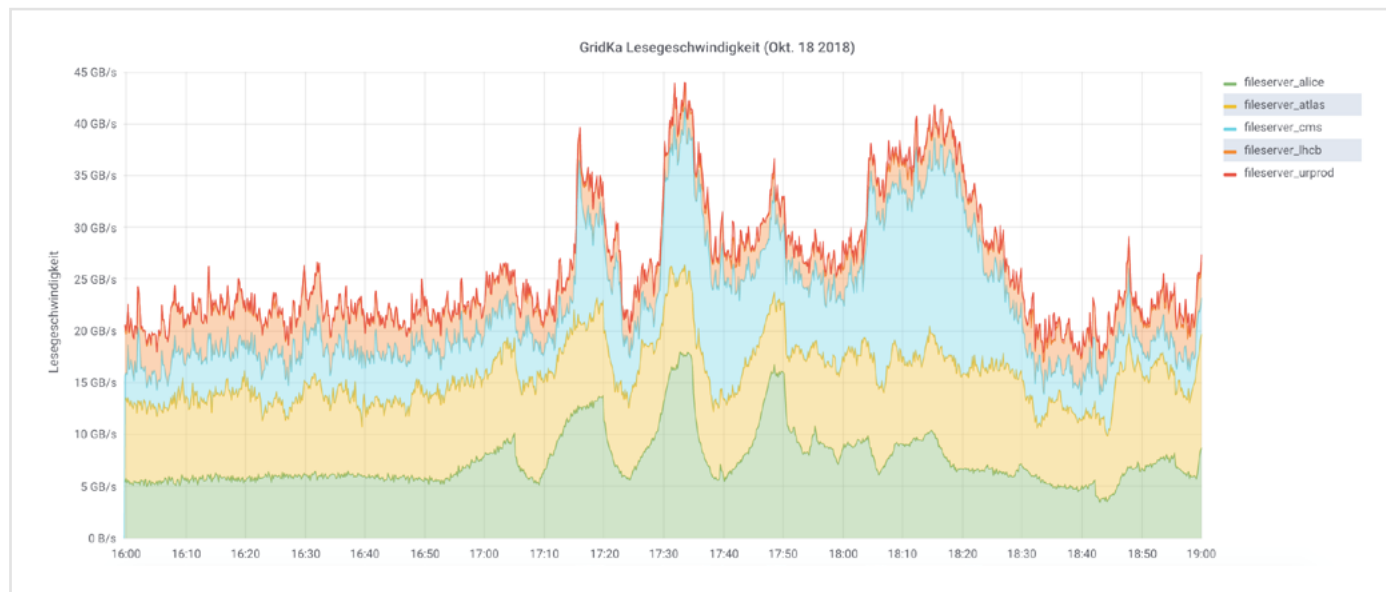


Abb. 2: Beispiel für ein Grafana-Dashboard. Dargestellt wird die totale Lesegeschwindigkeit des GridKa-Speichers gruppiert nach Experiment

### Technisches Design der Speichersysteme

Abbildung 1 zeigt schematisch den Aufbau der beiden Speichersysteme am Beispiel von GridKa. Dieses unterteilt sich in verschiedene Subcluster, die unterschiedliche Aufgaben erfüllen.

Der Storage-Cluster bildet und verwaltet die Platten-Pools und Dateisysteme. Nutzer des Systems können über mehrere Protokoll-Cluster Daten in die Dateisysteme schreiben oder lesen. Die Protokoll-Cluster mounten die Dateisysteme lediglich aus dem Storage-Cluster, sie verfügen aber über keine eigenen Speicherkapazitäten. In GridKa existiert typischerweise ein Protokoll-Cluster pro angeschlossenen Experiment. Die Unterteilung in Storage- und Protokoll-Cluster erlaubt eine organisatorische und

administrative Trennung des Speichers auf der einen und den notwendigen Zugriffsprotokollen auf der anderen Seite.

Performer Zugriff der Protokoll-Server auf die Daten mit hohem Durchsatz und niedriger Latenz wird durch Remote Direct Memory Access (RDMA) über zwei redundante 56 Gigabit/s FDR Infiniband-Fabrics erreicht. Die Infiniband-Fabrics sind in GridKa mit einem Blocking-Factor von 2:1 in einer Fat-Tree-Topologie mit jeweils 2 Spine- und 5 Leaf-Switches aufgebaut. In der LSDF existieren zwei 56 Gigabit/s FDR-Switches zur Anbindung aller Storage- und Protokoll-Server. Die Protokoll-Server sind jeweils über einen 40 Gigabit/s Ethernet-Link in GridKa bzw. zwei gebündelte Links mit 10 Gigabit/s in der LSDF angebunden.

GridKa	22 redundante Speicher-Server (NEC Express 5800 R120g-2M)	LSDF	6 redundante Speicher-Server (NEC Express 5800 R120g-2M)
	90 Disk-Enclosures (NetApp E5600 / DE6600)		29 Disk-Enclosures (NetApp E5600 / DE6600)
	5400 HDDs für Daten (8+10 TB)		1668 HDDs für Daten (8+10 TB)
	58 SSDs (1,6 TB) in separaten Enclosures für Dateisystem-Metadaten		54 SSDs (1,6 TB) in gemeinsamen Enclosures für Dateisystem-Metadaten
	Zwei redundante Infiniband-Fabrics (Mellanox 56 Gb/s FDR), Blocking-Faktor 2:1		Zwei redundante Infiniband-Fabrics (Mellanox 56 Gb FDR)
	64 Protokoll-Server mit jeweils 40 Gb/s Ethernet		20 Protokoll-Server mit jeweils 2x 10 Gb/s Ethernet
	8x 100 Gb Ethernet Uplink in das GridKa Backbone		

Tabelle 1: Technische Eigenschaften der Speichersysteme in der LSDF und in GridKa

Der Storage-Cluster besteht aus mehreren unabhängigen Storage Building Blocks (SBBs) und lässt sich daher einfach durch Hinzufügen weiterer Blöcke sowohl in seiner Größe als auch in seiner Leistung skalieren. Jeder SBB besteht aus 10 Enclosures, die in Fünferblöcken mit jeweils zwei Raid-Controllern redundant an zwei Server angeschlossen sind.

Jedes Enclosure hält 60 8 Terabyte oder 10 Terabyte Nearline-SAS-Festplatten. Mittels sogenannter Dynamic Disk Pools (DDP) [6] werden jeweils 50 Festplatten in logischen Disk-Pools zusammengefasst, in denen Daten in Raid6-Stripes (8+2P) geschrieben werden. In jedem Pool wird eine Speicherkapazität, die der Größe von drei Festplatten entspricht, reserviert und dient als Ersatz bei auftretenden Festplattendefekten. Verglichen mit klassischen Raid6-Systemen ermöglichen DDPs deutlich schnellere Rebuilds im Falle von Festplattenausfällen.

Der Storage-Cluster in GridKa enthält zwei zusätzliche Enclosures mit schnellem SSD-Speicher für die Metadaten der Spectrum Scale-Dateisysteme. Alle Metadaten werden in zwei Kopien geschrieben, um die Verfügbarkeit der Dateisysteme auch bei Ausfall eines der beiden Enclosures zu gewährleisten.

### Management und Monitoring

Alle Server werden mit Hilfe der Management-Tools Foreman und Puppet über eine zentrale Infrastruktur installiert und konfiguriert. Die notwendigen Puppet-Module wurden selbst

entwickelt und ermöglichen es, Server automatisch zu existierenden Spectrum Scale-Clustern hinzuzufügen. Da die Cluster-Konfiguration über viele Server mit mehreren Kopien verteilt gespeichert wird, sind alle Server im Wesentlichen zustandslos. Jeder beliebige Server kann so jederzeit automatisiert neu installiert werden.

Für das Monitoring werden momentan zwei Software-Stacks eingesetzt. Es wurde damit begonnen, Log-Dateien in einer Elasticsearch-NoSQL-Datenbank [7] zu sammeln und mit zusätzlichen Attributen anzureichern. Logfiles können entweder in einem Kibana- oder einem Grafana-Dashboard [8] durchsucht und visualisiert werden. Leistungsmetriken werden mit Telegraf ermittelt und in einer InfluxDB-Zeitreihendatenbank [9] gespeichert. Die Metriken können über ein Grafana-Dashboard zusammengefasst und visualisiert werden [10]. Abbildung 2 zeigt die aggregierte RDMA-Lese- und Schreibgeschwindigkeit für alle Server des GridKa-Speichersystems gruppiert nach Experiment.

Die neuen Speichersysteme in GridKa und der LSDF sind seit mittlerweile über einem Jahr erfolgreich in Produktion. Beide Systeme konnten in der Zwischenzeit im laufenden Betrieb erweitert werden. Zusätzliche reguläre Erweiterungen von mindestens 20% pro Jahr sind momentan in Vorbereitung.

### Weitere Informationen

[1] LSDF: [www.scc.kit.edu/forschung/11843.php](http://www.scc.kit.edu/forschung/11843.php)

[2] GridKa: [www.gridka.de](http://www.gridka.de)

[4] NEC GxFS: [www.nec.com/en/global/solutions/hpc/storage/gxfs.html](http://www.nec.com/en/global/solutions/hpc/storage/gxfs.html)

[5] IBM Spectrum Scale: [www.ibm.com/de-de/marketplace/scale-out-file-and-object-storage](http://www.ibm.com/de-de/marketplace/scale-out-file-and-object-storage)

[6] DDP: [www.netapp.com/us/info/what-is-dynamic-disk-pools-technology.aspx](http://www.netapp.com/us/info/what-is-dynamic-disk-pools-technology.aspx)

[7] ELK: [www.elastic.co/elk-stack](http://www.elastic.co/elk-stack)

[8] Grafana: [grafana.com](http://grafana.com)

[9] Telegraf / InfluxDB: [www.influxdata.com](http://www.influxdata.com)

[10] SDM Grafana: [grafana-sdm.scc.kit.edu](http://grafana-sdm.scc.kit.edu)

[11] [www.scc.kit.edu/dienste/11228.php](http://www.scc.kit.edu/dienste/11228.php)

### New storage systems for GridKa and the LSDF

The department Scientific Data Management in SCC currently operates two multi-petabyte storage systems with a total capacity of 44 petabytes. They are located in the Large Scale Data Facility and the Grid Computing Center Karlsruhe (GridKa) and provide central and reliable storage resources for large research data for users of different data-intensive scientific disciplines in the state of Baden-Württemberg and the Helmholtz Association as well as particle and astroparticle physics experiments supported by GridKa.

## Das neue SCC-Ticketsystem – erweiterte Möglichkeiten der Kommunikation

Für die Erfassung und Bearbeitung von Kundenanfragen hat das SCC ein neues Ticketsystem in Betrieb genommen. Dieses ersetzt schrittweise das bereits vorhandene Ticketsystem. Die neue Lösung bietet Kunden und Ticketbearbeitern eine Weboberfläche auf dem aktuellen Stand der Technik sowie zahlreiche neue Funktionalitäten.

Günter Grein

Für die Erfassung und Bearbeitung von Kundenanfragen aller Art (Service-Aufträge, Meldung von Störungen und Problemen etc.) nutzt das SCC schon lange ein sogenanntes Ticketsystem. In einem Ticketsystem wird für jede Kundenanfrage ein Vorgang (Ticket) mit einer eindeutigen ID angelegt. Sämtliche zur Vorgangsbearbeitung gehörende Kommunikation wird in diesem Ticket erfasst. Damit erleichtert ein Ticketsystem einerseits die Kommunikation zwischen Dienstbetreibern und Kunden und ermöglicht andererseits eine strukturierte Bearbeitung aller Kundenanfragen.

Bisher nutzte das SCC als Ticketsystem das „Materna Helpdesk“, eine Anwendung, die vom IT-Dienstleister Materna<sup>1</sup> erstellt wurde. Es basierte auf einem BMC Remedy<sup>2</sup> System. Diese Anwendung ist mittlerweile annähernd 20 Jahre alt. Sie hat im Laufe der Jahre viele, durch das SCC und seine Vorgängereinrichtungen selbst entwickelte, Änderungen und Anpassungen erfahren. Eine Weiterentwicklung des Systems und die Anpassung an den aktuellen technischen Standard wäre nicht oder nur mit großem Aufwand möglich. Daher war es an der Zeit das bisherige

### The new SCC ticket system - extended communication possibilities

The SCC has put a new ticket system into operation for recording and processing customer requests. This will gradually replace the existing ticket system. The new solution offers a state-of-the-art web interface for customers and ticket processors as well as numerous new functionalities.

Ticketsystem durch ein neues, moderneres System zu ersetzen.

### Standardkomponenten vereinfachen den Betrieb

Ein erklärtes Ziel des SCC ist es, ein Standardsystem zu nutzen und auf Eigenentwicklungen weitestgehend zu verzichten. Dadurch soll der Wartungs- und Pflegeaufwand möglichst gering gehalten werden. Als Produkt für das neue SCC-Ticketsystem wurde nach einer intensiven Evaluation Open-Source-Tool OTRS<sup>3</sup> ausgewählt, ein Produkt, das an zahlreichen Universitäten und Forschungseinrichtungen sowohl in Deutschland als auch weltweit im Einsatz ist.

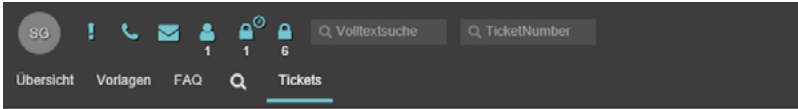
Der Zugang zum neuen SCC-Ticketsystem ist mit allen gängigen Webbrowsers möglich. Das System nutzt den zentralen Authentifizierungsdienst des KIT<sup>4</sup> und ermöglicht damit die Anmeldung mit dem persönlichen KIT-Benutzerkonto (Account). Somit entfällt eine separate Registrierung und Verwaltung von Nutzerkonten in OTRS.

Für die Zuordnung der SCC-Mitarbeiter zu den jeweiligen Arbeitsgruppen, die Tickets bearbeiten, greift OTRS auf eine Gruppenstruktur im KIT-weiten Verzeichnisdienst Active Directory<sup>5</sup> zurück. Über

die zentrale Gruppenverwaltung kann jeder Dienstbetreiber die Mitglieder seiner Arbeitsgruppen organisieren und ist nicht auf die Intervention der OTRS-Systemadministratoren angewiesen.

### Zugang für Beschäftigte und Studierende

In Q2/2019 ist geplant, den Zugang zum Ticketsystem für alle Beschäftigte und Studierende freizugeben. So können diese Nutzergruppen neben den bereits vorhandenen Kommunikationswegen Telefon, E-Mail und SCC Service Desk vor Ort nun ihre Anliegen auch bequem über ein Webformular an das SCC kommunizieren.



The screenshot shows a web interface for the SCC ticket system. At the top, there is a navigation bar with icons for home, help, and search, and a search bar with the text 'Volltextsuche' and 'TicketNumber'. Below the navigation bar, there are tabs for 'Übersicht', 'Vorlagen', 'FAQ', and 'Tickets'. The 'Tickets' tab is active, and the view is set to 'Ansicht nach Status: offenen Tickets'. The main content area shows a table of open tickets with columns for 'offenen Tickets 8' and 'Geschlossene Tickets 138'. The table has columns for 'Sammelaktion', 'Freigeben', 'Inhalt', 'Historie', 'Priorität', 'Notiz', 'Schließen', 'Verschieben', and 'Spam'. The table contains several rows of ticket data, including ticket numbers, creation times, and titles.

offenen Tickets 8		Geschlossene Tickets 138					
<input type="checkbox"/>		SCC#	▼ALTER	GEÄNDERT	TITEL		STATUS
<input type="checkbox"/>		6403290	4 h 18 m	19.02.2019 10:21	Einrichtung einer KIT-Teamseite		warten auf schließen
<input checked="" type="checkbox"/>		6403282	21 h 17 m	19.02.2019 10:34	HECTOR School Anfrage für neuen SharePoint		warten zur Erinnerung
<input type="checkbox"/>		6403279	23 h 18 m	18.02.2019 18:35	Löschung einer KIT-Teamseite		neu
<input type="checkbox"/>		6403272	1 d 0 h	18.02.2019 16:13	Dienst +KIT +Teamseiten		warten auf schließen
<input type="checkbox"/>		6403258	1 d 4 h	18.02.2019 16:14	Kontingent für KIT-Teamseite erhöhen		warten auf schließen

Abb. 1: Ticketübersicht für den Service Support.

Sobald sich das neue Ticketsystem in einem stabilen Regelbetrieb befindet, wird der Zugang für alle Beschäftigte und Studierende über das Self-Service-Portal my.scc.kit.edu freigegeben.

<sup>1</sup> www.materna.com/DE/Home/home\_node.html

<sup>2</sup> www.bmc.com/it-solutions/remedy-itsm.html

<sup>3</sup> otrs.com/de/home/

<sup>4</sup> www.scc.kit.edu/dienste/6921.php

<sup>5</sup> www.scc.kit.edu/dienste/8709.php



## Smart studieren am KIT = mobil studieren

SCC macht Studierende zu Studienbeginn mobil: Campus Management, ILIAS, KIT-Mail und KIT Mobile bieten „smarten“ Zugriff auf Studieninhalte und -administration.

Karin Schäufele



Großes Interesse an den mobilen Portalen (KIT Erstsemesterbegrüßung 2018)

Zum Wintersemester 2018/19 begrüßte das SCC in der Orientierungsphase mehr als 670 interessierte Studienanfänger, die sich in 18 Vorträgen über das SCC informierten. Vorgestellt wurden z.B. die Dienste Print und Plot, bwSync&Share, E-Mail, Sicherheitszertifikate, Netzzugang sowie die SCC-Poolräume. Wertvolle Hinweise zur Account-Verwaltung, Zugänge zu den Portalen Studierendenverwaltung und E-Learning wurden ergänzt durch die Bekanntgabe der Supportstellen ServiceDesk und TechPoint. Im Anschluss besuchten die Erstsemester in 32 Führungsgruppen die Rechnerräume des SCC. Hier wurde für alle erfahrbar, dass die Verarbeitung und Speicherung von Daten immer an raum- und energieintensive Hardware gebunden ist. Das SCC präsentiert sich so schon zu Beginn des Studiums als interessanter Forschungs- und Arbeitsplatz für angehende Studierende der Informatik, Mathematik und Ingenieurwissenschaften. Auch bei der KIT-Erstsemesterbegrüßung in der Karlsruher Schwarzwaldhalle nutzten die neuen Studierenden und ihre Eltern die Gelegenheit, Fragen zu stellen und Nützliches über das SCC zu erfahren.

Nach der „Generation Y“ hat jetzt die „i-Generation“ die Universitäten erreicht. Die Jahrgänge der Jahrtausendwende, die die amerikanische Persönlichkeitspsychologin Jean Twenge<sup>1</sup> auch als „Selfies“ bezeichnet, sind echte „Digital Natives“. Sie verwenden Laptops, Tablets und vor allem Smartphones so selbstverständlich wie die Generationen vor ihnen Papier und Bleistift. Verständlicherweise erwarten sie daher von ihrer Universität digitale Angebote, um das Studium auch mobil zu organisieren.



<sup>1</sup> Dr. Jean M. Twenge „Me, My Selfie and I: Was Jugendliche heute wirklich bewegt“, Mosaik Verlag, erschienen 29.05.2018

Das SCC stellt als Informationstechnologie-Zentrum des KIT sowohl Lehr- und Lernplattformen (ILIAS), Studienverwaltungsportale (Campus Management) als auch den mobilen Zugriff auf das KIT E-Mail-Postfach bereit. Just in time bietet das Portal KIT Mobile aktuelle Informationen z.B. Abfahrtszeiten des KIT-Shuttles, Mensa- und Casino-Speisepläne und, für Studierende besonders interessant, Vorlesungen im nahen Umkreis. Die angezeigte Gebäude-Nummer, Raumbezeichnung sowie Beginn und Ende leiten direkt zur Vorlesung. Darüber hinaus kann der ILIAS-Arbeitsbereich der Vorlesung über den Link in der Anzeige online geöffnet werden. KIT Mobile ist somit die „up to date“-Informationsplattform, auch für die spontane Organisation des Studientages.

Die Web-Auftritte der o. g. Portale sind für die Smartphone-Nutzung optimiert. Die sichere Anmeldung zu persönlichen


Bereichen wie z.B. ILIAS-Schreibtisch oder Prüfungsan- und abmeldung erfolgt über das zentrale Single-Sign-On-System des KIT. Werden die Webseiten im gleichen Browser geöffnet, müssen KIT-Konto und Passwort nur einmal eingegeben werden. Die Konfiguration von WLAN, VPN und Mail ist mit einigen wenigen Schritten erledigt und wird auf den SCC-Seiten ausführlich beschrieben. Insbesondere die einfache Einbindung des KIT-Exchangekontos in das Smartphone-interne Mail-Feature hilft den Studierenden, relevante Mitteilungen z.B. von Dozenten, Studienverwaltung oder SCC nicht zu übersehen. Werden die Webseiten als Verknüpfungen auf dem Startbildschirm des Smartphones abgelegt, ist der Zugang zu den Angeboten komfortabel und direkt möglich. Natürlich profitieren neben den Studierenden auch Mitarbeiterinnen und Mitarbeiter sowie Gäste und Partner des KIT von den mobilen Angeboten.

### Smart study at KIT = mobile study

SCC makes students mobile at the beginning of their studies: Campus Management, ILIAS, KIT-Mail and KIT Mobile offer "smart" access to study content and administration.


SCC welcomed more than 670 interested first-year students in the orientation phase for winter semester 2018/19, who informed themselves about the SCC in 18 lectures. For example, the services Print and Plot, bwSync&Share, e-mail, security certificates, network access and the SCC pool rooms were presented. Valuable information on account management, access to the student administration and e-learning portals were supplemented by the announcement of the ServiceDesk and TechPoint support offices. The first semester students then visited the computer rooms of the SCC in 32 guided groups. Here it became clear to everyone that the processing and storage of data is always tied to space- and energy-intensive hardware. SCC thus presents itself as an interesting research and workplace for prospective computer scientists, mathematicians and engineers right from the start of the studies. SCC was also present at the regular first year welcome event of KIT in the Schwarzwaldhalle in Karlsruhe to offer new students and their parents an opportunity to ask questions and learn useful things about SCC.

As the information technology centre of KIT, SCC offers teaching and learning platforms (ILIAS), study administration portals (campus management) and mobile access to the KIT mailbox. Just in time, the KIT Mobile portal provides up-to-date information, e.g. departure times of the KIT shuttle, canteen and casino menus and, particularly interesting for students, lectures in the immediate vicinity. The displayed building number, room name as well as start and end lead directly to the lecture. In addition, the ILIAS work area of the lecture can be opened online via link in the advertisement. KIT Mobile is so an up to date information platform for the spontaneous organisation of the study day.




KIT  
Karlsruher Institut für Technologie


### Mobil im KIT\*




**Aktuell und informativ**  
m.kit.edu



**Rund ums Studium**  
campus.studium.kit.edu



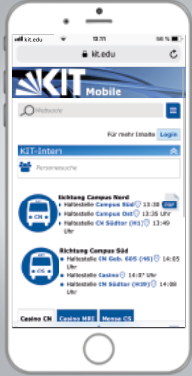
**Lehren und Lernen**  
ilias.studium.kit.edu



**KIT Mails to go**  
scc.kit.edu/dienste/mail2go.php

\*keine App nötig

STEINBUCH CENTRE FOR COMPUTING



KIT - Die Forschungsuniversität in der Helmholtz-Gemeinschaft www.kit.edu

Just in time Informationen unter m.kit.edu



Foto: Markus Breig

## IT-Angriffe: GEMEINSAM das KIT schützen

Um Schäden für das KIT und seine Mitglieder und Angehörigen zu vermeiden, sind IT-Sicherheitsvorfälle umgehend dem KIT-CERT zu melden. Meldepflichtig sind der Einbruch von Hackern, die Verbreitung oder Ausführung von Schadcode, der Kontrollverlust über Zugangsdaten oder andere wichtige Daten – auch durch den Verlust von Datenträgern oder Geräten. Andere Vorgänge können freiwillig gemeldet werden.

*Andreas Lorenz*

Im Rahmen des aktualisierten IT-Sicherheitskonzeptes des KIT gibt es nun eine explizite Meldepflicht bei IT-Sicherheitsvorfällen. Doch was genau heißt das? Was muss gemeldet werden?

Die IT-Sicherheit ist nicht die Computerpolizei, niemand soll verhaftet werden. Sie will auch nicht verhindern oder blockieren, sondern unterstützen, helfen, Lösungen finden und somit Nutzen bringen. Wenn jemand etwas versehentlich nicht erkennt oder nicht meldet ist das nicht strafbar. Denn man kann nur etwas melden, das man auch erkannt hat.

Melden Sie IT-Sicherheitsvorfälle bitte umgehend an Ihren IT-Beauftragten oder schicken Sie eine E-Mail an [cert@kit.edu](mailto:cert@kit.edu). Gemeinsam mit Ihnen analysieren und besprechen wir, was getan werden kann, um das Risiko so gering wie möglich zu halten.

Eine Meldepflicht besteht u.a. bei folgenden IT-Sicherheitsvorfällen wie beispielsweise dem Verlust von Geräten (z.B. PCs, Laptops, Smartphones), über die Sie auf Dienste oder Daten des KIT

zugreifen, oder von Datenträgern (z.B. USB-Sticks, CDs), auf denen z.B. Passwörter, Klausuren, Bewerbungen, Noten oder Gehaltsabrechnungen gespeichert sind. Informieren Sie ebenfalls, wenn Sie eine E-Mail oder Nachricht mit betrügerischem Inhalt (z.B. eine Phishing-E-Mail) nicht als solche erkannt, sondern z.B. auf den Link geklickt oder den Anhang geöffnet haben sowie beim Entdecken von Schadsoftware auf Geräten. Ein Hinweis für Schadsoftware ist, dass Geräte sich plötzlich anders verhalten (z.B. sehr langsam werden).

Manchmal tauchen in den Diensträumen plötzlich Geräte auf, wie z.B. WLAN-Router, kleine Boxen, andere PCs/Laptops, wurden aber nicht angekündigt – auch das kann ein Sicherheitsvorfall darstellen. Sprechen Sie uns auch an bei Erpressungsversuchen oder Nötigung, sich nicht regelkonform zu verhalten, z.B. wenn jemand Unbekanntes unbedingten Zugriff auf Ihre Geräte oder Ihre Räume haben möchte.

Weitere Informationen finden Sie unter [www.scc.kit.edu/sid/meldepflicht](http://www.scc.kit.edu/sid/meldepflicht)



## Produktivsetzung des Anlageninventursystems im Universitätsbereich

Zur Durchführung der Anlageninventur wird seit 2005 am Großforschungsbereich ein Barcode-gestütztes Inventursystem eingesetzt. Seit Oktober 2018 ist das maschinell unterstützte Verfahren im Universitätsbereich etabliert und im SAP-System integriert. Ein weiterer Schritt zur Harmonisierung der Geschäftsprozesse am KIT.

*Gerald Helck, Jürgen Stehmer<sup>1</sup>*

Das Verfahren einer maschinell unterstützten Anlageninventur sieht vor, zunächst die Anlagen mit maschinenlesbaren Inventaretiketten zu versehen. Im Rahmen der eigentlichen Inventurmaßnahme werden diese dann über mobile Lesegeräte gescannt und das Inventurergebnis via SAP-Schnittstelle im Anlagenstammsatz hinterlegt. Der Einsatz eines Anlageninventursystems führt zu einer erheblichen Beschleunigung des Erhebungsprozesses. Aufgrund des einfachen Verfahrens treten Erfassungsfehler nur noch selten auf, so dass die Nacharbeiten seitens der Anlagenbuchhaltung deutlich reduziert werden.

Für neu beschaffte Anlagen werden die Etiketten im Zusammenhang mit der Zugangsbuchung der Anlage erzeugt. Des Weiteren erfolgt auch der Ausdruck relevanter Informationen zur Anlage (z.B. Anlagennummer, Bezeichnung, Lieferant, Nutzungsdauer, Projektnummer) in der Anlagenkarte, einem Formular, das geeignet ist, den Anlagenbestand der jeweiligen Organisationseinheit zu katalogisieren und vorzuhalten. Beides, Anlagenkarte und Anlagenetikett, werden zur Weiterverwendung an die Organisationseinheit versandt.

FIMA/Anlagenbuchhaltung hatte das SCC beauftragt, die Nutzung des Anlageninventursystems auf den Universitätsbereich auszudehnen. Hierzu mussten zusätzliche Nutzungsrechte der Software erworben sowie die bestehende Hardware-Infrastruktur um drei Barcodelesegeräte sowie

einen Barcodedrucker erweitert werden. Die Besonderheit der dezentralen Wareneingangsbuchung am Universitätsbereich über das SAP-SRM hatte zur Folge, dass auch die Programme für Etiketten- bzw. Anlagenkartendruck angepasst werden mussten. Ein Ausdruck der Etiketten und Karten direkt über die Zugangsbuchung des zentralen Wareneingangs, wie am Großforschungsbereich praktiziert, ist hier nicht möglich. Deshalb wurden die Programme so erweitert, dass Etiketten und Karten für alle im benutzerdefinierten Zeitraum zugegangenen Anlagen über einen „Massendruck“ ausgegeben werden. Um einen möglichst effizienten Versand der Papiere zu erreichen, können

Etiketten und Karten nun nach vom Fachbereich frei vorgebbaren Kriterien sortiert werden. Ferner wurden die Programme mit einem Simulationsmodus versehen, um vor dem eigentlichen Druck das anstehende Volumen abschätzen und Fehlersituationen beheben zu können.

Das neue Verfahren ist seit 01.10.2018 produktiv und läuft reibungslos. Die Anlageninventur ist nach den Vorgaben der Landeshaushaltsordnung (LHO) durchzuführen. Im konkreten Fall nimmt das betreffende Institut Kontakt mit der Anlagenbuchhaltung auf. Vor der eigentlichen Inventur muss der Anlagenaltbestand mit Barcodeaufklebern versehen werden.



Scanner und Etikettendrucker für die Anlageninventur

<sup>1</sup> Finanzmanagement (FIMA) des KIT



# Neuerungen von A-Z: Die Lernplattform -> einmal komplett umgekrempelt

Die Lernplattform des KIT hat umfassende technische Neuerungen erfahren. Seit September 2018 laufen die Webserver unter der aktuellen Debian-Version „stretch“ und sind nun auch über IPv6 erreichbar. Zudem wurde die Lernplattform auf die neueste PHP-Version 7.x umgestellt.

Franziska Wandelmaier

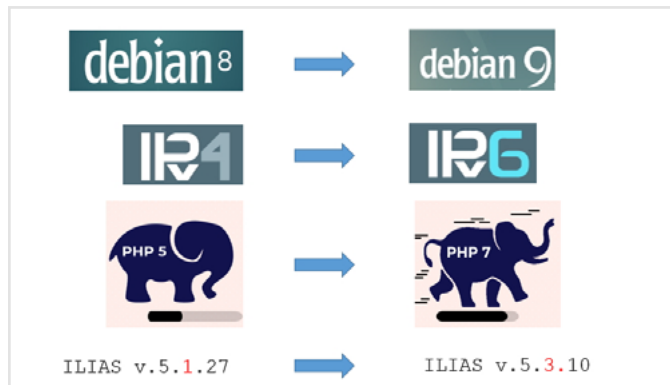


Abb. 1: Optimierungs- / Verbesserungsmaßnahmen

Die Lernplattform des KIT hat ein neues Zuhause gefunden: Am 12. September 2018 sind die vier ILIAS-Webserver in ein neues, sichereres VLAN gezogen und laufen seither unter der aktuellen Debian Version „stretch“. Sie sind nun auch über IPv6 erreichbar. Stretch brachte die neueste PHP-Version 7.x mit, auf welche die Lernplattform kurz nach ihrem eigenen Versions-Update umgestellt wurde (Abb. 1). Die Lernplattform hat sich nämlich nicht nur eine neue Wohlfühlumgebung gesucht, sondern auch einer Verjüngungskur unterzogen: Der Sprung von ILIAS v5.1.27 nach v.5.3.10 brachte nicht nur das schnellere PHP mit sich, sondern hat die Nutzung der Lernplattform an einigen Stellen interessanter und attraktiver gemacht. Hierzu ein paar Beispiele:

## Teilnehmerverwaltung

- Kursadministratoren können nun ausgewählte Kursmitglieder gezielt zusammenfassen und diese als Mitglieder einem anderen Kurs oder einer Gruppe hinzufügen.
- Teilnehmerlisten können individuell zusammengestellt und anschließend ausgedruckt werden.

## Interaktion fördern/aufzeigen

- Eine Timeline (Abb. 2) ermöglicht es, die Zusammenarbeit der Kursmitglieder attraktiv und als lebendiges Bild abzubilden. Hierzu kann man seinen ILIAS-Kurs so einstellen, dass neue Ereignisse, Inhalte und Nachrichten auf einer Zeitachse in einem eigenen Reiter „Timeline“ dargestellt werden.

- Portfolios können dank verbesserter Vorlagen nun viel leichter und schneller angelegt, gepflegt und über den Seiteneditor auch überall eingebunden werden: Kursbetreuer können so z.B. innerhalb ihrer Kurse auf ihr eigenes Portfolio verlinken, um den Teilnehmenden die Kontaktaufnahme zu erleichtern. Darüber hinaus können die erstellten Portfolios nun auch als PDF gespeichert oder ausgedruckt werden. Die Grundidee folgt dabei Thomas Häcker<sup>1</sup>: Es geht um die Darstellung der eigenen Entwicklung, des eigenen Könnens bzw. der eigenen Leistungen, wobei der/die Autor/in des eigenen Portfolios darüber entscheiden, was er oder sie von sich preisgeben möchte.
- Der komplett überarbeitete Kalender bietet viele neue Ansichten und Funktionen. Anfahrtsbeschreibungen und Kursmaterialien können jetzt für anstehende Termine gesammelt heruntergeladen werden.

## Motivation durch Belohnung

- Seit dieser Version können Dozierende „Badges“ an ihre Kursteilnehmer vergeben. Diese können Studierende ihrem eigenen Profil beifügen und dort speichern. Die Vergabe derartiger Abzeichen könnte also beispielsweise an das erfolgreiche Bearbeiten von Kursinhalten gekoppelt sein, genauso können sie manuell von Lehrenden, z.B. für gute Mitarbeit, im Forum oder in der Präsenzveranstaltung vergeben werden.

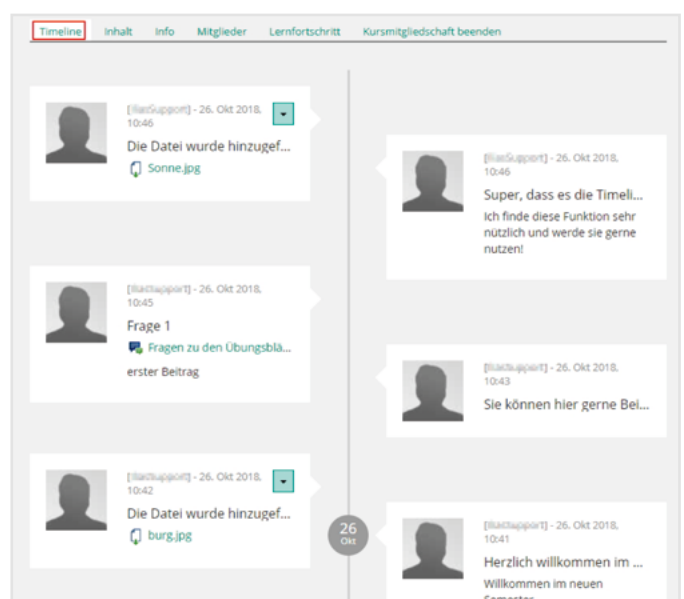


Abb. 2: Darstellung der Interaktion innerhalb eines Kurses

<sup>1</sup> de.wikipedia.org/wiki/Thomas\_H%C3%A4cker

## Der Helmholtz-Inkubator Information & Data Science

Datenbasierte Forschung entwickelt sich in atemberaubendem Tempo. Daher wurde der Helmholtz-Inkubator gegründet, um die Kompetenzen der Gemeinschaft auf diesem Gebiet zu bündeln und informationsbasierte Forschung völlig neu zu definieren.

*Andreas Kosmider<sup>1</sup>, Achim Streit*

Wissenschaft und Forschung erzeugen heute enorme Datenmengen und eröffnen damit gänzlich neue Perspektiven für den Erkenntnis- und Informationsgewinn. Die Verarbeitung und Analyse dieser komplexen und immer umfangreicheren Datenmengen ist für das gesamte Wissenschaftssystem eine der größten Herausforderungen der Zukunft. Helmholtz ist auf diesem Gebiet mit seinem weiten Spektrum von Big Data Analytics, Supercomputing, dem gesamten Data-Lifecycle sowie der Softwareentwicklung bis hin zu künstlicher Intelligenz und Robotik hervorragend positioniert. In allen Einzeldisziplinen, aber gerade auch in ihrem Zusammenspiel, ergeben sich vielfältige Möglichkeiten.

Der Helmholtz-Inkubator, 2016 vom Helmholtz-Präsidenten Wiestler ins Leben gerufen, verfolgt das Ziel, kreative Köpfe aus der gesamten Gemeinschaft regelmäßig in Interaktion zu bringen, Grundlagen für innovative, interdisziplinäre Netzwerke und Ansätze zu schaffen sowie zukunftsweisende Themenfelder und neuartige Pilotprojekte zu identifizieren. So katalysierte der Helmholtz-Inkubator die Entwicklung von fach- und disziplinübergreifenden Projektvorhaben, welche die üblichen Disziplin- und Forschungsbereichsgrenzen überwinden. Im Sommer 2017 wurden die ersten fünf interdisziplinären Pilotprojekte ausgewählt – darunter das Helmholtz Analytics Framework, HAF (siehe Vorstellung mit ersten Ergebnissen in dieser Ausgabe, S. 31), gemeinsam koordiniert vom Jülich Supercomputing Centre (JSC) und dem SCC. Parallel dazu wurden in einem zweijährigen Prozess und unter Beteiligung von mehr als 140 Wissenschaftlerinnen und Wissenschaftlern und zehn Beraterinnen und

Beratern aus namhaften, forschenden Unternehmen folgende Themenfelder mit strategischer Bedeutung für die gesamte Helmholtz-Gemeinschaft identifiziert:

- Helmholtz-Plattform für Künstliche Intelligenz und Maschinelles Lernen (Helmholtz Artificial Intelligence Cooperation Unit, HAICU)
- Helmholtz-Plattform für Bilddatentechnologien (Helmholtz Imaging Platform, HIP)
- Helmholtz-Plattform für Basistechnologien und grundlegende Dienste für datenbasierte Großforschung (Helmholtz Infrastructure for Federated ICT Services, HIFIS)
- Helmholtz-Plattform für wissenschaftlichen Nachwuchs und Graduiertenschulen als vernetzter Verbund von fünf regionalen Helmholtz Information & Data Science Schools (HIDSS) und der Helmholtz Information & Data Science Academy (HIDA)

Im Inkubator und den oben genannten Themenfeldern ist das SCC erfolgreich engagiert. Das SCC trug maßgeblich zur Konzeptionierung der Helmholtz Information & Data Science Academy (HIDA) bei, mit der ein Helmholtz-weiter Verbund zur postgradualen Aus- und Weiterbildung in Information & Data Science geschaffen wird. Ziel ist, ein breites Wissen zu Methoden und Technologien von Information & Data Science und deren Anwendung konsequent in den Forschungsthemen der Helmholtz zu verankern. An einer regionalen Graduiertenschule ist das SCC ebenfalls beteiligt: In der HIDSS4Health (Helmholtz Information & Data Science School for Health) werden das KIT, das Deutsche Krebsforschungszentrum (DKFZ) und die

Universität Heidelberg gemeinsam junge Forscherinnen und Forscher ausbilden, die an der Schnittstelle von Gesundheitsforschung, Lebens- und Datenwissenschaften arbeiten. In diesen Verbund aus HIDSS und HIDA bringt das SCC eine langjährige Erfahrung in der Organisation und Durchführung der GridKa School sowie Expertise in der Forschung, Lehre und Weiterbildung zu Daten-intensivem Hochleistungsrechnen und föderiertem Datenmanagement groß-skaliger Forschungsdaten ein.

Weiterhin ist das SCC am Aufbau einer Helmholtz-Infrastruktur für föderierte IT Services (HIFIS) beteiligt. HIFIS hat zum Ziel, eine nahtlose, Zentren-übergreifende IT-Infrastruktur mit integrierten Dienstleistungen auf der Basis schneller Netze und einem einheitlichen Nutzerzugang (AAI) zu etablieren, einen in die Zusammenarbeits- und Forschungsprozesse integrierten, sicheren, effizienten und weltweit verfügbaren Daten- und Anwendungszugriff auf Basis von Cloud-Diensten zu realisieren sowie Ausbildung und Unterstützung in der Entwicklung hochwertiger und nachhaltiger Software zu leisten. HIFIS knüpft dazu unter anderem an die Helmholtz-Data Federation (HDF, koordiniert vom KIT) an und befördert den Ansatz föderierter IT-Infrastrukturen. Das SCC wird seine langjährige Erfahrung in der Entwicklung und Bereitstellung föderierter Identitätsmanagement-Infrastrukturen (u.a. bWDM, AARC) sowie föderierter Daten- und Rechen-Infrastrukturen (u.a. WLCG, GridKa, HDF, EOSC, bwCloud, HNSciCloud, INDIGO-DataCloud, DEEP-HybridDataCloud) maßgeblich einbringen.

<sup>1</sup> Helmholtz-Geschäftsstelle, Bereichsleiter Strategische Initiativen

Darüber hinaus hat das SCC in der Arbeitsgruppe zur Formulierung der Helmholtz-Plattform für Künstliche Intelligenz und Maschinelles Lernen (HAICU) mitgearbeitet. Ein fünftes Themenfeld zum Thema Metadaten wird derzeit noch weiter detailliert. Auch hier ist das SCC maßgeblich beteiligt.

#### Weitere Informationen

[www.helmholtz.de/aktuell/presseinformationen/artikel/artikeldetail/helmholtz\\_investiert\\_kuenftig\\_zusaetzliche\\_35\\_millionen\\_euro\\_jaehrlich\\_in\\_die\\_digitalisierung\\_der\\_forsch/](http://www.helmholtz.de/aktuell/presseinformationen/artikel/artikeldetail/helmholtz_investiert_kuenftig_zusaetzliche_35_millionen_euro_jaehrlich_in_die_digitalisierung_der_forsch/)

[www.helmholtz.de/forschung/information\\_data\\_science/helmholtz\\_inkubator/](http://www.helmholtz.de/forschung/information_data_science/helmholtz_inkubator/)

[kit.edu/kit/pi\\_2018\\_121\\_datenschatze-fur-die-medizin-nutzbar-machen.php](http://kit.edu/kit/pi_2018_121_datenschatze-fur-die-medizin-nutzbar-machen.php)

## Inkubator-Workshop am 29. und 30.10.2018 in Berlin



*Helmholtz-Präsident Wiestler eröffnet den Workshop.*

*Kanzleramtsminister Braun beantwortet Fragen des Auditoriums.*

### The Helmholtz Incubator for Information & Data Science

The Helmholtz president has formed the Helmholtz incubator for information & data science to identify and strengthen specific topics of information & data science with a high strategic importance to the whole Helmholtz Association. These are artificial intelligence and machine learning, imaging procedures & analysis methods, basic ICT technologies and services for large-scale data-driven research as well as training of young scientists at newly designed graduate schools within a national consortium. SCC is significantly involved in the Helmholtz Information & Data Science Academy (HIDA) that aims at establishing knowledge about methods and technologies of information & data science and their application across all research fields of the Helmholtz Association and acts as an umbrella for the newly established regional graduate schools. SCC is also involved in the regional graduate school HIDSS4Health (Helmholtz Information & Data Science School for Health) focusing on interdisciplinary PhD researchers in health, life and data science. Furthermore, SCC is heavily involved in HIFIS (Helmholtz Infrastructure for Federated ICT Services) to build up a Helmholtz-wide AAI infrastructure and cloud services for data-intensive research, contributed to the concept of HAICU (Helmholtz Artificial Intelligence Cooperation Unit) and is very active in further detailing the metadata topic.

## Nachhaltigkeitsaspekte bei der Entwicklung der Ginkgo-Software-Bibliothek

Oft beginnt ein Softwareprojekt mit einer Promotions- oder Masterarbeit: Man schreibt das erste Softwareprojekt, ohne die Prinzipien nachhaltiger Softwareentwicklung zu berücksichtigen wie zum Beispiel Dokumentation oder Qualitätschecks. Alles wird dem Ziel untergeordnet, einen Code zum Laufen zu bekommen, um die Promotions- oder Masterarbeit erfolgreich abschließen zu können. Denn der Erfolg der wissenschaftlichen Arbeit hängt von den entwickelten Algorithmen, also den Features ab, und nicht von der Nachhaltigkeit der Implementierung. Am Beispiel der Softwarebibliothek Ginkgo zeigen Wissenschaftler der Forschungsgruppe FiNE des SCC, wie HPC-Algorithmen nachhaltig implementiert und getestet werden.

Hartwig Anzt



Ginkgo [1] ist eine Open Source C++ Library für Lineare Algebra, die den sogenannten SOLID Prinzipien folgt [2]: Diese sind Single responsibility, Open/closed principle, Liskov substitution, Interface segregation und Dependency principle. Mit dem Ziel, die Kompatibilität und die Erweiterbarkeit zu maximieren, werden in Ginkgo die Algorithmen von den architekturenspezifischen Implementierungen separiert. Diese strikte Trennung und die gleichzeitige Verwendung eines „Executors“ (ähnlich einem Ticket für eine bestimmte Zugverbindung) ermöglicht es, Soft- und Hardware Backends hinzuzufügen oder an die Hardwareentwicklung anzupassen. Derzeit unterstützt Ginkgo die Parallelisierung mit OpenMP sowie die Grafikprozessorplattform CUDA von NVIDIA. Zusätzlich gibt es einen sequentiellen Reference-Executor, der in Komponententests (unit tests) die parallelen Implementierungen auf Korrektheit überprüft.

<sup>1</sup> [github.com/ginkgo-project/ginkgo](https://github.com/ginkgo-project/ginkgo)

<sup>2</sup> [en.wikipedia.org/wiki/SOLID](https://en.wikipedia.org/wiki/SOLID)

<sup>3</sup> [github.com/abseil/googletest](https://github.com/abseil/googletest)

<sup>4</sup> [en.wikipedia.org/wiki/Doxygen](https://en.wikipedia.org/wiki/Doxygen)

<sup>5</sup> [en.wikipedia.org/wiki/BSD\\_licenses](https://en.wikipedia.org/wiki/BSD_licenses)

Das Ginkgo-Softwareprojekt beginnt nicht mit der Realisierung von Features, sondern mit dem Design der Softwarebibliothek und der Implementierung eines nachhaltigen Entwicklungsprozesses.

[ginkgo-project.github.io/](https://ginkgo-project.github.io/)

Jedes Feature der Ginkgo-Bibliothek muss einen Komponententest besitzen, der von Google Test [3] orchestriert wird. In Ginkgo werden nicht nur Basiskonzepte der Linearen Algebra als lineare Operatoren betrachtet, wie zum Beispiel das Matrix-Vektor-Produkt, sondern auch komplexe Algorithmen, wie iterative Löser und Vorkonditionierer. Die Klasse der Linearen Operatoren (LinOp) vererbt zentrale Funktionalitäten wie „apply“, „clone“, „copy“ sowie die jeweiligen Realisierungen. Das reduziert den Implementierungsaufwand und erhöht die Nutzerfreundlichkeit. Um den Einstieg in Ginkgo zu vereinfachen, setzt das Entwicklerteam zudem auf die Dokumentation mit dem Werkzeug Doxygen [4], umfangreiche Beispielcodes und Tutorials.

Im Sinne eines nachhaltigen Entwicklungsprozesses ist Ginkgo als Open Source Code unter BSD-3 [5] lizenziert. Voraussetzung zur Integration in das zentrale Git-Repository sind zwei Code Reviews. Der Master-Zweig wird automatisch in ein nicht-öffentliches Git-Repository gespiegelt, um neue Ideen und noch nicht veröffentlichte Algorithmen zu schützen - auch dies gehört zur akademischen Softwareentwicklung. Die Synchronisation der Master-Zweige im öffentlichen und privaten Repository ermöglicht die rasche Integration erfolgreich neu entwickelter Algorithmen im Produktionscode.

men zu schützen - auch dies gehört zur akademischen Softwareentwicklung. Die Synchronisation der Master-Zweige im öffentlichen und privaten Repository ermöglicht die rasche Integration erfolgreich neu entwickelter Algorithmen im Produktionscode.



Ginkgo Core-Team: Hartwig Anzt, Terry



Sowohl das öffentliche Repository als auch die Entwicklungszweige sind in ein kontinuierliches Integrationssystem (Continuous Integration, CI [6]) eingebettet. Dieses testet mit CMake [7], ob der Code auf den unterschiedlichen Hardware-Architekturen und unterschiedlichen Compiler-Umgebungen lauffähig ist und die Komponententests erfolgreich abschließen (Cross-Platform Portability [8]). Neue Features, die die CI-Anforderungen erfüllen und von zwei Reviewern bewilligt wurden, werden in den Master-Zweig der Bibliothek integriert und deren Leistung auf einem HPC-System getestet.

Für Entwickler, die keine Zugriffsrechte auf ein HPC-System haben, stellt dieser Entwicklungszyklus einen Anreiz dar, solche Performancetests auch für ihre Softwarebeiträge zu erstellen. Da dies aber per se ein Sicherheitsrisiko darstellt, müssen die Reviewer externer Beiträge die volle Verantwortung für die Integrität der Module übernehmen.

Die Laufzeitergebnisse neuer Features werden vom HPC-System in ein öffentliches Archiv kopiert und können mit einer Webanwendung, dem Ginkgo Performance Explorer (GPE, [9]), interak-

tiv analysiert werden. Das ermöglicht es nicht nur die Laufzeit einzelner Routinen mit Ginkgo über die Entwicklungszeit zu verfolgen, sondern erlaubt auch externen Entwicklern, die Performance der beigetragenen Features mit Hilfe individueller Jsonata-Skripte zu analysieren.

Die Forschungsgruppe FiNE ist davon überzeugt, dass dieser Software-Entwicklungszyklus wesentlich dazu beiträgt, dass sich Ginkgo als prominente Softwarebibliothek etabliert.

**Weitere Informationen zum Entwicklungszyklus der Ginkgo-Software unter: [icl.utk.edu/~hanzt/tmp/PerformanceDatabase.pdf](http://icl.utk.edu/~hanzt/tmp/PerformanceDatabase.pdf)**

<sup>6</sup> en.wikipedia.org/wiki/Continuous\_integration

<sup>7</sup> cmake.org/

<sup>8</sup> en.wikipedia.org/wiki/Software\_portability

<sup>9</sup> ginkgo-project.github.io/gpe/



Cojean, Goran Flegar, Thomas Gruetzmacher und Pratik Nayak (v. l.).

# The Art of Writing Scientific Software in an Academic Environment

Hartwig Anzt

It all starts with a PhD project. Or a Master's project. We write our first piece of scientific software, not in a sophisticated software design with appropriate documentation or correctness tests, but with the single goal of having the code running and delivering the results or insights we need to graduate. Sometimes, we later come back to that code, remembering that some years ago we had programmed something that may now be useful. And we try to reconstruct and re-engineer what we had once started, but then abandoned as we moved on. Even though we learn the hard way the advantages of following strict software design rules and having a well-designed software development process, we often do not adopt it in our first professional appointment in academia; the focus is on publishing papers, getting community attention, and maybe writing proposals – but not on sustainable software development. The code may get more complex, and we probably start using versioning systems for collaborating with peers and employ some correctness checks, but the goal remains the same: to realize new ideas and algorithms as quickly as possible – we feel our future depends more on the publication of our next paper than on the usefulness and sustainability of the related code stack.

For me personally, the situation changed when the first external users started using my software. Suddenly, questions like “where is that documented?” and “are these two components compatible?” arise. This is the point when we start regretting not having written proper documentation in the first place – and the point when we realize how hard it is to remember ideas and considerations from half a year ago. Moving on, we try to introduce tests that check the correctness of our routines and cross-platform compatibility – an adventure doomed to

fail as the code typically neither adheres to the single responsibility principle (SRP) nor the segregation of the interface (ISP) [1]. In response to the situation, we design tests that ensure the correctness of some key functionality and commonly used execution patterns. While this may provide a glimpse of what is working, it can never serve as a bulletproof test for all scenarios.

Moving forward, the only option is often to rewrite everything from scratch. At that point, it is important to make a high-level analysis of the efforts:

- Does the software have long-term potential, justifying a complete rewrite?
- Do I have the resources (personnel/funding) to realize a redesign?
- Can I survive academically a period with few/no scientific publications or project proposal submissions?
- Do I have the infrastructure, support, and expertise to realize a redesign in a healthy software development workflow?

Aside from these critical questions determining the feasibility of a redesign effort, one has to consider aspects concerning the existing software stack: Will it be maintained in the future? Is backward-compatibility an option? How can existing users be served during the software redesign?

Personally, I have the privilege of several funding efforts and significant community support, which allowed me to meet all requirements.

In the Ginkgo [2] software effort we turn the development upside down: we do not start with implementing features, we start with the design of the library structure and a healthy software development process.

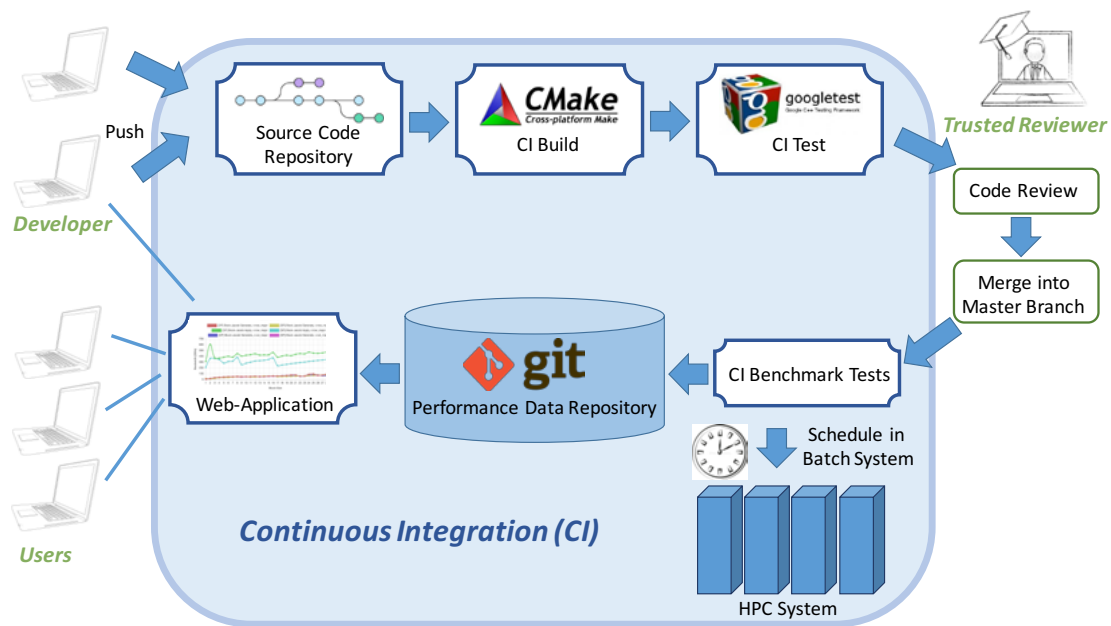
Ginkgo is designed as an open-source C++ linear algebra library following the SOLID software design principles [1]: Single responsibility; Open/closed; Liskov substitution; Interface segregation and the Dependency principle. With the goal of maximizing compatibility and extensibility, we decided to completely separate the linear algebra algorithms from the architecture-specific kernel implementations. Using an architecture-specific “Executor” allows adding, removing, or modifying backends according to future changes in the hardware architectures and parallelization strategies. Currently, Ginkgo is designed for node parallelism, featuring backends for NVIDIA GPUs and OpenMP-supporting platforms. Additionally, it features a sequential reference executor that is used in the unit tests to ensure the correctness of the hardware-specific parallel kernels. The unit tests are realized using the Google Test [3] framework, and having new features covered by unit tests is a prerequisite for merging them into the master branch of the repository. Another central design feature of Ginkgo is the concept of expressing not only basic linear algebra operations like matrix-vector products but also complex algorithms such as iterative solvers and preconditioners as linear operators that all share key functionalities like apply, clone, copy, etc. This greatly enhances user friendliness, reduces the coding effort, and efficiently avoids code duplication. To simplify its use, library features are not only thoroughly documented using Doxygen [4], but are typically also accompanied by small usage examples or tutorials.

<sup>1</sup> [github.com/ginkgo-project/ginkgo](https://github.com/ginkgo-project/ginkgo)

<sup>2</sup> [en.wikipedia.org/wiki/SOLID](https://en.wikipedia.org/wiki/SOLID)

<sup>3</sup> [github.com/abseil/googletest](https://github.com/abseil/googletest)

<sup>4</sup> [en.wikipedia.org/wiki/Doxygen](https://en.wikipedia.org/wiki/Doxygen)



Über diesen Entwicklungsprozess, hier schematisch dargestellt, entstehen Softwaremodule mit hoher Qualität und Zuverlässigkeit.

To enable a healthy software development process, Ginkgo is a community effort licensed under the modified BSD license [5]. A central Git repository requires two reviews on every merge to the master branch. This branch is automatically mirrored into a private repository where branches are used for the development of novel algorithms and the deployment of unpublished performance optimizations. In this way, Ginkgo keeps in mind academic software development, where the option of keeping new ideas and algorithms confidential is important in avoiding knowledge theft and in adhering to the principles of academic publication. Synchronization with the public master branch enables quick integration of successful development into production code.

Public and private feature developments are automated with the help of a continuous integration (CI) [6] framework that checks the CMake [7] compilation process for a large number of hardware architectures and compiler/library environments (cross-platform portability [8]). The CI system also ensures successful completion of the unit tests. If all tests pass successfully and two reviewers approve the merge request, a new feature is integrated into Ginkgo's master branch and benchmarked on an HPC system. While this also allows external contributors to have their features executed on an HPC cluster (which they may not even have access to), the reviewers approving the merge take full responsibility for the integrity of the code. The performance results for the new feature are then retrieved from the

HPC system, and archived in a repository storing performance data of Ginkgo routines. The Ginkgo performance explorer (GPE) [9] is a web application that can retrieve the performance data from the repository and visualize it in user-defined fashion, with the help of the integrated JSONata scripting language. This not only enables monitoring performance changes over time, but also allows external contributors to receive feedback about the feature they contributed – and it can potentially optimize the implementation. We are convinced that this software development cycle, along with the sophisticated library design, forms the basis for a successful academic community effort and allows Ginkgo to transition from a "small community" project to a high-profile software ecosystem.

**Further reading about the software development cycle we implement for Ginkgo at:**  
[icl.utk.edu/~hanzt/tmp/PerformanceDatabase.pdf](http://icl.utk.edu/~hanzt/tmp/PerformanceDatabase.pdf)  
[bssw.io/blog\\_posts/the-art-of-writing-scientific-software-in-an-academic-environment](http://bssw.io/blog_posts/the-art-of-writing-scientific-software-in-an-academic-environment)

<sup>5</sup> en.wikipedia.org/wiki/BSD\_licenses

<sup>6</sup> en.wikipedia.org/wiki/Continuous\_integration

<sup>7</sup> cmake.org/

<sup>8</sup> en.wikipedia.org/wiki/Software\_portability

<sup>9</sup> ginkgo-project.github.io/gpe/

## Effiziente Rechenlastverteilung für die Simulation von turbulenter Verbrennung auf Supercomputern

Numerische Simulationen von Verbrennungsprozessen werden meistens mit Vereinfachungen der zugrundeliegenden physikalischen und chemischen Prozesse durchgeführt, um die Rechenanforderungen gering zu halten. Will man aber Flammen mit einem hohen Detailgrad simulieren, müssen diese Prozesse durch genaue Modelle berücksichtigt werden. Solche detaillierten Modelle umfassen beispielsweise die Bildung und den Abbau von bis zu tausend Zwischenprodukten, die bei der Verbrennung entstehen können. Zudem muss die Rechnung eine mikrometeregenaue Auflösung haben, um die Struktur der Flammen erfassen zu können. Detaillierte numerische Simulationen sind daher nur auf Supercomputern möglich. Mitarbeiter des SimLab Energy am SCC haben in Kooperation mit dem Engler-Bunte-Institut (EBI) nun eine Methode vorgestellt, die die parallele Berechnung von Verbrennungsvorgängen auf Supercomputern beschleunigt. Die Arbeit „Optimizing Load Balancing of Reacting Flow Solvers in OpenFOAM for High Performance Computing“ wurde auf der OpenFOAM Konferenz in Hamburg mit dem Best Paper Award ausgezeichnet.

*Thorsten Zirwes*

Auch heute noch wird ein Großteil des Energiebedarfs der Welt durch Verbrennungsprozesse gedeckt. Es ist daher notwendig, diese Prozesse im Hinblick auf Effizienz, Brennstoffbedarf und Schadstoffemissionen zu optimieren. Allerdings gibt es noch viele offene Fragen über die komplizierte Wechselwirkung zwischen chemischen Reaktionen und Strömungsphänomenen.

Oft ist es schwer, genaue Messungen von Verbrennungsvorgängen vorzunehmen, da viele Vorgänge, wie zum Beispiel chemische Reaktionen, sehr schnell ablaufen. Zudem finden technisch relevante Verbrennungsprozesse oft bei hohen Temperaturen und hohen Drücken statt, was Messungen zusätzlich erschwert. Eine Alternative zu aufwändigen Experimenten sind daher numerische Simulationen. Will man hochaufgelöste und detaillierte Ergebnisse von Flammen erhalten, können die Simulationen aufgrund der hohen Komplexität von Verbrennungsvorgängen nur auf modernen Supercomputern durchgeführt werden, wie sie beispielsweise am SCC betrieben werden. Das zu berechnende Gebiet wird dabei in viele kleine Teilgebiete unterteilt und die mathematischen Gleichungen in

den Teilgebieten parallel gelöst. So können zehntausende CPU-Kerne gleichzeitig an der Simulation arbeiten.

Ein Problem, das dabei auftauchen kann, ist, dass die Arbeit nicht gleichmäßig auf die CPU-Kerne verteilt wird. Im schlimmsten Fall braucht ein CPU-Kern viel länger, um sein Teilgebiet zu berechnen, als alle anderen. Die Folge ist, dass die anderen CPU-Kerne warten müssen, bis der langsame aufgeholt hat. Durch das Warten werden so teure Supercomputer-Ressourcen verschwendet.

Bei der Berechnung von Verbrennungsvorgängen tritt diese Art der ungleichmäßigen Arbeitsverteilung häufig bei der Berechnung der chemischen Reaktionsgeschwindigkeiten auf. Diese Berechnung benötigt oft einen Großteil der Gesamtrechnenzeit, da selbst für einfache Brennstoffe wie Methan viele hundert verschiedene chemische Reaktionen berücksichtigt werden müssen. In Abb. 1 links ist ein Foto einer Methan-Bunsenflamme dargestellt. Man erkennt in Hellblau leuchtend die Bereiche, in denen der Hauptteil der chemischen Verbrennungsreaktionen abläuft. In der Mitte von Abb. 1 ist ein Ausschnitt aus der numerischen Simulation einer

ähnlichen Flamme dargestellt. Die Farben zeigen, wo chemische Reaktionen – hier die Reaktion von Methan – ablaufen. Typischerweise findet der Großteil der Reaktionen nur in einem dünnen Bereich, der Flammenfront, statt. Will man jetzt diese Simulation parallel durchführen, muss man das Rechengebiet in Teilbereiche zerlegen. Rechts in Abb. 1 ist ein Beispiel gezeigt, wie man diese Flamme in 12 Teilbereiche (Domains) unterteilen könnte. Das Problem wird schnell deutlich: In Domain 1 liegt ein Teil der Flamme, der besonders hohe Reaktionsraten hat (dargestellt in rot). Es muss daher besonders viel Rechenaufwand betrieben werden, um diese Reaktionsraten zu berechnen. In Domain 0 dagegen reagiert überhaupt nichts. Der Rechenaufwand in diesem Teilgebiet ist also viel geringer. In vielen relevanten Anwendungen kommen turbulente Flammen zum Einsatz. Diese bewegen sich unvorhersehbar durch das Rechengebiet, sodass man vor Durchführung der Simulation nicht weiß, in welchem Teilgebiet sich die Reaktionszone zu welcher Zeit aufhalten wird. Eine Verbesserung der Arbeitsverteilung muss also dynamisch während der Simulation passieren.



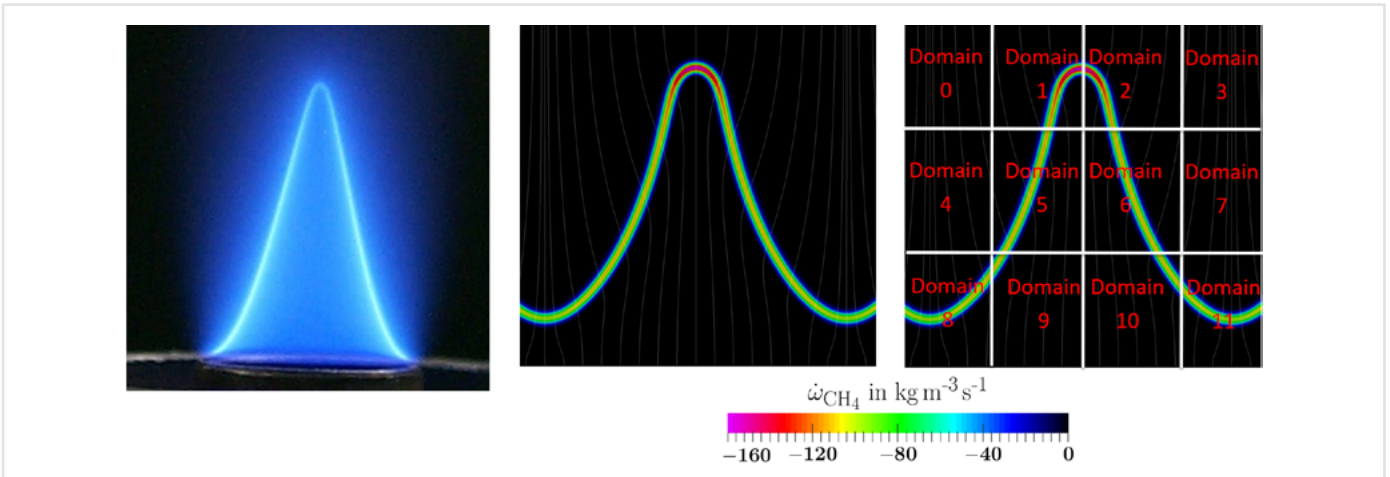


Abb. 1: Links: Foto einer Bunsenflamme. Mitte: Simulation einer Bunsenflamme. Die Farben zeigen, wo chemische Reaktionen stattfinden. Rechts: Ein Beispiel, wie die Simulation dieser Flamme in 12 Teilgebiete (Domains) aufgeteilt werden kann.

In der Arbeit „Optimizing Load Balancing of Reacting Flow Solvers in OpenFOAM for High Performance Computing“ wurde eine Methode präsentiert, um eine dynamische Lastverteilung während der Simulation zu erzeugen, die einige Besonderheiten der Berechnung chemischer Reaktionsgeschwindigkeiten ausnutzt.

mit hohem Rechenaufwand. Dies zeigt, dass die standardmäßig von OpenFOAM bereitgestellten Mittel zur Berechnung chemischer Reaktionsgeschwindigkeiten zu unterschiedlicher Lastverteilung und damit zu einer ineffizienten Nutzung der Supercomputer-Ressourcen führen.

Um diese Probleme zu umgehen, wurde ein Lastverteilungsverfahren in OpenFOAM implementiert, das speziell die Berechnung der chemischen Reaktionsraten optimiert. In dieser Methode werden Paare von Prozessen gebildet, wobei ein Prozess ein besonders schneller und der

Zunächst wurden Performance Messungen auf verschiedenen Supercomputern<sup>1</sup> durchgeführt, um die Lastverteilung genauer zu charakterisieren. Die Simulationen wurden mit dem open-source toolkit OpenFOAM durchgeführt, welches zur Simulation von reagierenden Strömungen benutzt werden kann. OpenFOAM bietet verschiedene Löser an, um numerische Strömungssimulationen mit chemischen Reaktionen durchzuführen. In Abb. 2 ist eine solche Messung mit 120 Prozessen dargestellt: Jede horizontale Zeile repräsentiert einen parallelen Prozess. Der Verlauf auf der x-Achse zeigt die Simulationszeit an. Grün und Grau sind Zeiten, in denen die Prozesse „nützliche Arbeit“ durchführen. In Rot markiert sind Zeiten, zu denen der Prozess wartet. Im besten Fall sollten also niemals rote Bereiche zu sehen sein. Man erkennt, dass es nur wenige Prozesse gibt, die relativ lange für ihre Berechnung brauchen (wenig Rot in der Zeile). Die meisten Prozesse allerdings warten auf die wenigen Prozesse

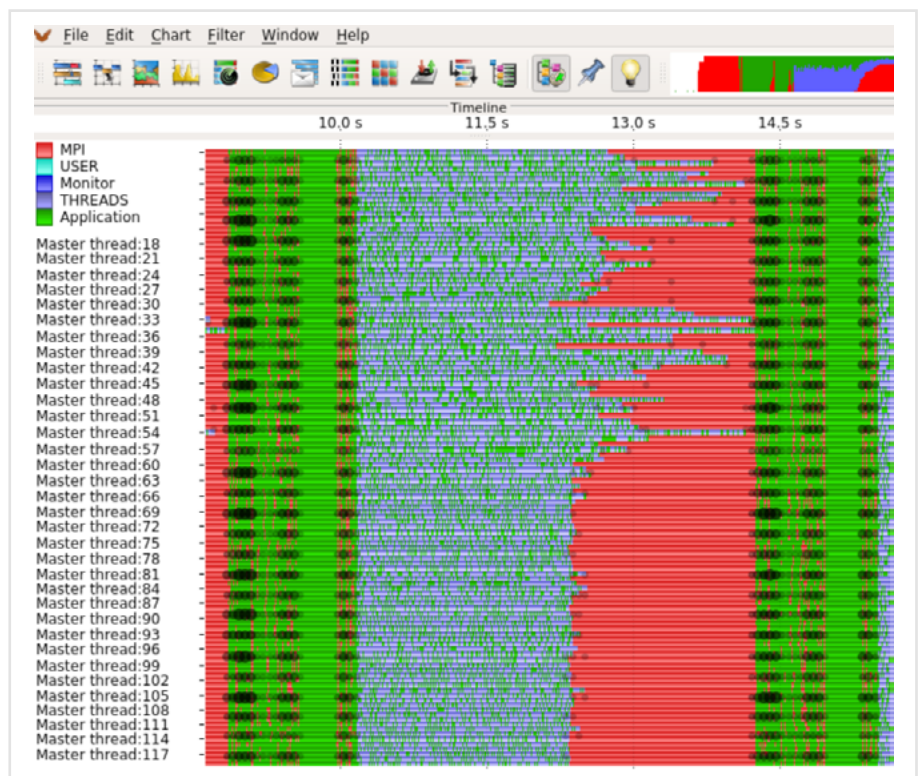


Abb. 2: Performance-Messung vor der Lastverteilungs-Optimierung: Jede horizontale Zeile repräsentiert einen parallelen Prozess. Der Verlauf auf der x-Achse zeigt die Simulationszeit an. Grün und Grau sind Zeiten, in denen der Prozess nützliche Arbeit durchführt. In Rot markiert sind Zeiten, zu denen der Prozess warten muss.

<sup>1</sup> ForHLR II am SCC, Rechencluster am HLRS

andere ein besonders langsamer Prozess ist. Diese beiden Prozesse schieben dann während der Simulation einen Teil ihrer Arbeit dynamisch hin und her. Dabei wird ausgenutzt, dass chemische Reaktionsgeschwindigkeiten aus Temperatur, Konzentrationen und Druck an einem Punkt ausgerechnet werden können, ohne die Werte in der Umgebung berücksichtigen zu müssen. Dadurch wird es ermöglicht,

die anfallende Arbeit sehr frei zu verteilen. Das Resultat ist, dass Prozesse, die eigentlich sehr viel Zeit benötigen würden, ihre Arbeit mit einem Prozess teilen, der nicht viel zu tun hat. Dies reduziert besonders lange Rechenzeiten einzelner Prozesse, sodass alle Prozesse insgesamt weniger warten müssen. So kann die Rechenzeit in vielen Fällen um 30 % reduziert werden.

In Abb. 3 ist das Temperaturfeld einer Flamme dargestellt, die auf dem Supercomputer Hazel Hen am Höchstleistungs-Rechenzentrum Stuttgart mit über 28.000 Prozessen berechnet wurde. Die Anwendung der im oben genannten Paper vorgestellten Methode zur verbesserten Lastverteilung führt dazu, dass eine Million Kernstunden bei der Rechnung eingespart werden können.

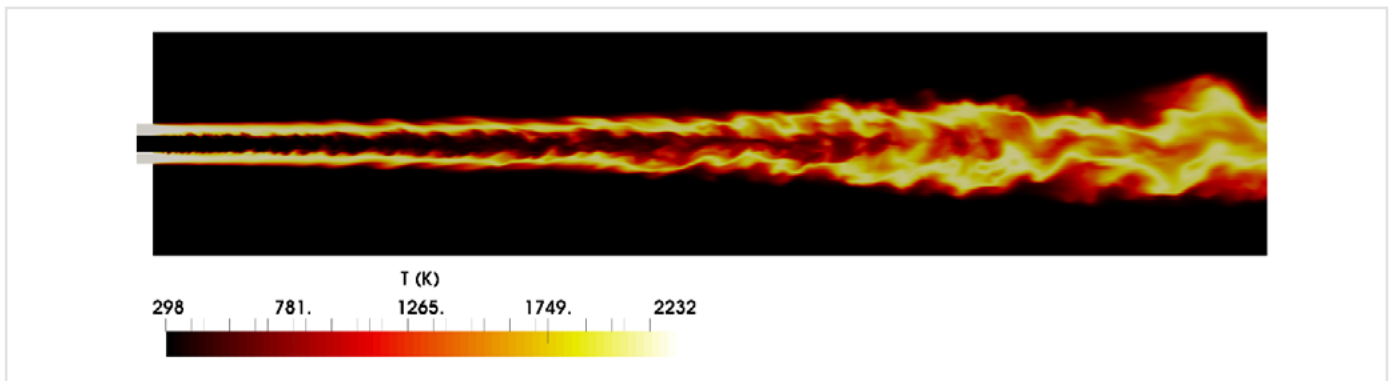


Abb. 3: Temperaturfeld auf einem 2D Schnitt durch die Flamme.

### Improved Load Balancing for the Simulation of Flames on Supercomputers

Numerical simulations of combustion processes are usually carried out with many simplifications in order to keep the computing effort low. However, if flames are to be simulated with a high degree of detail, many physical processes must be taken into account. Such detailed simulations are only possible on supercomputers. The paper "Optimizing Load Balancing of Reacting Flow Solvers in OpenFOAM for High Performance Computing" by the SimLab Energy at SCC in cooperation with the Engler-Bunte-Institute (EBI) presents a method that accelerates the parallel calculation of combustion processes on supercomputers. When calculating chemical reaction rates, it is often the case that some processes require more computing time than others. The basic idea, therefore, is to form pairs of processes that share their work of calculating chemical reaction rates. This is made possible due to the fact that the calculation of chemical reaction rates can be more freely distributed among the processes than the calculation of many other quantities. The methodology was implemented in OpenFOAM, which itself does not provide load balancing for the calculation of chemical reaction rates. The paper won the Best Paper Award at the 6th OpenFOAM Conference in Hamburg.



Abb. 4: Verleihung des Best Paper Awards an Thorsten Zirwes auf der "6th OpenFOAM Conference" in Hamburg.

# Helmholtz Analytics Framework

Das Pilotprojekt Helmholtz Analytics Framework (HAF), vom KIT und dem Forschungszentrum Jülich gemeinsam koordiniert, wird die Entwicklung der Datenwissenschaften in der Helmholtz-Gemeinschaft stärken. Gemeinsam mit vier weiteren Helmholtz-Zentren, DESY, DKFZ, DLR und HMGU, werden in einem Co-Design-Ansatz zwischen Domänenwissenschaftler und Datenanalyseexperten herausfordernde Anwendungsprobleme aus den jeweiligen Helmholtz-Zentren untersucht. Konkret handelt es sich dabei um Fragen zur Erdsystemmodellierung, Strukturbio-logie, Luft- und Raumfahrt, Neurowissenschaft und Medizinischen Bildgebung.

*Markus Götz, Björn Hagemeier<sup>1</sup>*

## HELMHOLTZ Analytics Framework

Das Helmholtz Analytics Framework (kurz: HAF) ist ein Pilotprojekt im Bereich Data Science und wird durch die Helmholtz-Gemeinschaft mit nahezu sechs Millionen Euro über 36 Monate gefördert. Insgesamt sind sechs Helmholtz Zentren – DESY, DKFZ, DLR, FZJ, HMGU und KIT<sup>2</sup> – an dem Vorhaben beteiligt, das sich mit der systematischen Entwicklung domänenspezifischer Datenanalysetechniken und -algorithmen beschäftigt, um das Portfolio Data Science und Scientific Big Data Analytics (SBDA) innerhalb der Helmholtz-Gemeinschaft auszubauen. Dabei kommt ein sogenannter Co-Design Ansatz zum Einsatz. Das bedeutet, dass Wissenschaftler aus den jeweiligen Domänen eng mit Experten der Informationsverarbeitung zusammenarbeiten, um einen effektiven und effizienten Gebrauch der Datenanalysemethoden zu ermöglichen. Als nicht mehr wegzudenkender Bestandteil der modernen Wissenschaft hilft Scientific Big Data Analytics heute schon bei der Beantwortung zentraler Forschungsfragen. Hohe Datenkomplexität und rechenintensive Verfahren erfordern dabei den Einsatz wegbereitender Technologien aus dem Datenmanagement und dem Hochleistungsrechnen, um Datenzugriff, -integration, -verarbeitung und -assimilation zu ermöglichen. Die dafür im HAF entwickelten Lösungen werden zunächst an den Anwendungsfällen des Projektes erprobt und dann als generische Dienste bzw. Werkzeuge für den allgemeinen Gebrauch verfügbar gemacht. HAF wird durch das Projekt Helmholtz Data Federation (HDF) ergänzt, ein weiteres Helmholtz-Inkubatorprojekt, das sich mit der Entwicklung einer föderierten Datenverarbeitungsinfrastruktur für die Helmholtz-Gemeinschaft beschäftigt. Die Anfang 2017 gestartete HDF wird mit 50 Millionen Euro gefördert und dient unter anderem als Plattform für die Erprobung der im HAF entwickelten Datenanalysemethoden und deren Implementierungen. Mit den Projekten HAF und HDF legt die Helmholtz-Gemeinschaft ein solides Fundament für die immer stärker datengetriebene Wissenschaft der Zukunft (s. Artikel Seite 22).

## Anwendungsfälle

In insgesamt acht Anwendungsfällen aus fünf Forschungsbereichen werden wissenschaftliche Fragestellungen zu den Themen Erdsystemmodellierung, Neurologie, Luft- und Raumfahrt, Photonexperimente und Strukturbio-logie untersucht.

## Erdsystemmodellierung

Im Anwendungsfall Erdbeobachtung und Vorhersagen wird versucht, den Wasser- und Energiekreislauf unseres Planeten möglichst exakt in die Zukunft zu projizieren. Die Hauptschwierigkeit dieser Aufgabe liegt darin, die schiere Komplexität des Systems und die Anzahl der zu berücksichtigenden Einflussvariablen beherrschbar zu halten. Gelingt dieses Vorhaben, können sozioökonomisch relevante Vorgänge wie z.B. Fluten, Dürren oder auch die Folgen des Klimawandels besser verstanden und vorhergesagt werden. Die wichtigsten zum Einsatz kommenden SBDA-Methoden sind Ensemble Datenassimilationsverfahren und genetische Algorithmen.

Cloud and Solar Power Prediction ist der Titel des zweiten Anwendungsfalls, in dem versucht wird, die Bewegung von Wolken exakt vorherzusagen. Diese können einen nicht zu vernachlässigenden Effekt auf die Photochemie der Atmosphäre haben und damit, durch Schattenverdeckung, auch auf die Stromproduktion mit Solaranlagen. Eine kurzfristige Bewegungsvorhersage der Wolken auf Basis von Satellitenbildern ist ein bisher ungelöstes Problem und wird im Rahmen von HAF mit parallelen und skalierbaren Klassifikationsverfahren wie z.B. kaskadierenden Support Vector Machines untersucht.

Der dritte und letzte Anwendungsfall im Themenkomplex Erdsystemmodellierung beschäftigt sich mit dem Einfluss der Stratosphärenchemie und -physik auf Wetter und Klima am Boden. Die HAF-Forscher setzen dabei auf einen Hindcast-Ansatz. Das bedeutet, dass sie ein Ensemble aus leicht unterschiedlichen Simulationen der Atmosphäre durchführen und dabei in der Vergangenheit (1960er Jahre) beginnen und bis zum heutigen Tag rechnen. Dabei entstehen immense Datenmengen, die miteinander abgeglichen, die Simulation für zukünftige Vorhersagen verbessern. Mit Beginn des Projektes wird zunächst das einfacher

<sup>1</sup> Forschungszentrum Jülich - Jülich Supercomputing Centre (JSC)

<sup>2</sup> Deutsches Elektronen-Synchrotron (DESY), Deutsches Krebsforschungszentrum (DKFZ), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Forschungszentrum Jülich (FZJ), Helmholtz Zentrum München – Deutsches Forschungszentrum für Gesundheit und Umwelt (HMGU), Karlsruher Institut für Technologie (KIT)



zu bestimmende Klima auf der Südhalbkugel studiert und im Projektverlauf auf die Nordhalbkugel gewechselt.

### Neurologie

Die moderne medizinische Forschung, vor allem in Bereichen wie der personalisierten Medizin und dem Gehirnverständnis, beschäftigt sich mit der Fragestellung, wie sich Genetik und Umwelt auf beobachtete Phänotypen (z.B. Struktur, Funktion oder Veränderung über die Zeit) von gesundem und krankem Gewebe auswirken. Vielfach werden dafür bildgebende Verfahren eingesetzt, die dann von Algorithmen des maschinellen Sehens ausgewertet werden. Die HAF-Forscher aus dem Anwendungsfall High-Throughput Image-Based Cohort Phenotyping verwenden dafür vor allem Clusteringverfahren und Deep Learning. Heutige Bilder sind so hoch aufgelöst, dass die Auswertung eines einzelnen Bildes in der Regel mit Großrechnern erfolgen muss. Im Anwendungsfall Multi-Scale Multi-Area Interaction in Cortical Networks wird die Funktionsweise des Gehirns auf kleinstmöglicher Skala, den Neuronen, untersucht. Dazu werden aufgezeichnete Experimente ausgewertet, bei denen die Neuronenaktivität im visuellen und motorischen Teil des Gehirns während vorgegebener körperlicher Bewegungen gemessen wurde. Durch Mustererkennung sollen wiederholt zeitgleich oder in Reihe feuerverne Neuronen identifiziert werden, um auf ihre Funktion und ihr Verhalten rückzuschließen. Die große Herausforderung dabei ist, dass nur eine sehr kleine Zahl an Neuronen gemessen werden kann, diese aber schon eine solch große Datenmenge produzieren, dass auf parallele Datenverarbeitung zurückgegriffen werden muss. Sequenzanalyseverfahren und Korrelationsstudien eignen sich dafür besonders gut.

### Luft- und Raumfahrt

Die HAF-Kollegen aus dem DLR beschäftigen sich unter anderem mit der Erforschung neuer Designs moderner Zivilflugzeuge. Da dieser Vorgang sehr zeitintensiv und teuer ist, wird versucht, möglichst gute neue Kandidaten auf Basis von vereinfachten Modellen und Auswertung von Daten alter Designs vorab zu schätzen. Um den vielen aerodynamischen Nicht-Linearitäten der Strömungen und der Flugkurvenhülle gerecht zu werden, kommen im Anwendungsfall Virtual Aircraft parallele Algorithmen der hoch-dimensionalen Klassifikation, Surrogaten-Modellierung und Datenfusion zum Einsatz.

### Photonenexperimente

Im Anwendungsfall Automated Volumetric Interpretation werten HAF-Forscher am DESY Bilddaten aus, die in Röntgenstrahl-experimenten in Synchrotronen oder mit Freielektronenlasern gesammelt werden. Materialien oder kleine Lebewesen, wie z.B. spezielle Kristalle oder auch Bakterien, werden mit elektromagnetischer Strahlung beschossen, um Einblick in ihr Innerstes zu

erlangen und ihren Aufbau zu erkunden. Die dabei entstehenden Streuprofile sind nicht direkt verständlich und werden mit Hilfe von Segmentierungs-, Klassifikations- und Rekonstruktionsverfahren in ein Bild oder Volumen übersetzt. Die Qualität der Übersetzung hängt dabei direkt von der Menge der verfügbaren Eingabedaten ab: Je mehr desto besser. Um ein qualitativ hochwertiges Ergebnis zu erhalten, müssen die Datenanalyseverfahren also in der Lage sein, parallel große Datenmengen zu verarbeiten.

### Strukturbiologie

Die HAF-Forscher in der Strukturbiologie beschäftigen sich mit der Strukturbestimmung von Biomolekülen. Oft lässt sich daraus nämlich direkt ableiten, welche Funktion das Molekül innerhalb eines Biosystems hat. Experimentelle Ansätze, wie z.B. Röntgenstrahllexperimente, sind allein zumeist nicht ausreichend, um die komplexen zeitlichen und räumlichen Strukturanordnungen zu verstehen. Deshalb werden heutzutage Daten aus unterschiedlichen Experimenten zusammengeführt und mit Hilfe von Simulationen und Datenanalyseverfahren ausgewertet. Im HAF werden NMR Spektroskopie, Cryo-Elektronenmikroskopie und Molekulardynamiksimulationen verwendet. Als algorithmische Hilfsmittel verwenden die Strukturbiologen vor allem Bayes'sche Parametersuchen, Mustererkennungsverfahren wie selbst-organisierende Karten sowie Neuronale Netzwerke für die Strukturvorhersage.

### Aktueller Stand

Nachdem das Projekt nun bereits seit Januar 2018 läuft, hat es einige spannende Entwicklungen gegeben. Im Bereich der Anwendungsforschung sind außergewöhnliche, domänenübergreifende Forschungskollaborationen entstanden. So tauschen sich beispielsweise Neurologinnen mit Strukturbiologen und Forschenden der Erdsystemmodellierung über die Verwendung des Sequenzanalyseverfahren ASSET aus, während Luft- und Raumfahrer etwas von Klimaforscherinnen über Bayes'sche Parametersuchen lernen. Aber auch innerhalb wissenschaftlicher Felder bereitet das HAF den Weg für die Arbeit an neuen Forschungsfragen. In der Strukturbiologie zum Beispiel haben Physikerinnen und Physiker am KIT, die computergestützte Analysen von Proteinen durchführen, erstmalig Zugriff auf experimentelle Daten des Helmholtz-Zentrums in München. Bei den generischen Methoden hat sich das HAF zunächst mit einer Klassifizierung der bereits eingesetzten und im Rahmen der Anwendungsfälle aktuell untersuchten Datenanalyseverfahren beschäftigt. Dadurch konnte ein Anforderungskatalog für die HAF-Auswertungssoftware und -Tools erstellt werden. Nach einer umfangreichen Designphase ist auf dieser Basis HeAT (Helmholtz Analytics Toolkit) konzipiert worden. HeAT<sup>1</sup> ist eine Open-Source Programmierbibliothek, die Algorithmen für die Auswertung von Daten auf Hochleistungsrechnersystemen anbietet. Die in der Programmiersprache Python geschriebene Software erlaubt nicht



nur die transparente Benutzung von Beschleunigerhardware wie Grafikkarten (GPGPUs), sondern abstrahiert auch von der expliziten Programmierung von Kommunikationscode durch Anwender. Die HeAT wird dabei durch die o.g. Anwendungs-fälle getrieben. So wird der Sequenzanalysealgorithmus von mehreren eingesetzt und werden soll, als erstes Demonstrationsprojekt mit HeAT realisiert.



grammierung  
tionscode durch  
Entwicklung von  
durch die o.g.  
fälle getrieben.  
Sequenzana-  
ASSET, der  
Projektpartnern

### Danksagung

Das Projekt wird durch den Impuls- und Vernetzungsfond der Helmholtz-Gemeinschaft gefördert (Projektnummer: ZT-I-0003)

### Zitate



Es gibt eine Vielzahl von Datenanalyse-Tools, aber keine bedient die Nische High-Performance Computing ausreichend. HeAT könnte diese Lücke schließen mit seinem innovativen Ansatz der Mischung von verteiltem und hybridem Rechnen.

*(Kai Krajssek, Jülich Supercomputing Centre (JSC), Forschungszentrum Jülich (FZJ), HAF-Mitglied WP2)*



Eine solche Kollaboration zwischen Experimentalforschern und uns wäre ohne HAF nicht zustande gekommen.

*(Alexander Schug, Steinbuch Centre for Computing (SCC), Karlsruher Institute für Technologie (KIT), HAF-Mitglied WP1, Anwendungsfall Strukturbilogie)*



All-hands Meeting des HAF, 14.08.2018, DLR Köln

© DLR

### The Helmholtz Analytics Framework (HAF)

The Helmholtz Analytics Framework (HAF) pilot project, jointly coordinated by KIT and the Research Center Jülich, aims at fostering the data science development in the Helmholtz Association. Together with four other Helmholtz centers, namely DESY, DKFZ, DLR and HMGU, a systematic co-design approach between domain scientists and information experts is employed in order to work on challenging domain use cases of the respective partners. Concretely, the HAF research attempts to get deeper insight into the fields of Earth system modelling, structural biology, neurosciences, medical imagery as well as aeronautics and aerospace.

<sup>1</sup> [github.com/helmholtz-analytics/heat/](https://github.com/helmholtz-analytics/heat/)



## KIT-Zentrum Mathematics in Sciences, Engineering, and Economics (MathSEE) hat seine Arbeit aufgenommen

Das neu gegründete KIT-Zentrum "Mathematics in Sciences, Engineering, and Economics" (MathSEE) bündelt die interdisziplinäre mathematische Forschung am KIT. Es ist daher zwischen Theorie und Anwendung sowie zwischen Universitäts- und Großforschung angesiedelt. Forscher aus Mathematik, Naturwissenschaften sowie Ingenieur- und Wirtschaftswissenschaften arbeiten in interdisziplinären Forschungsprojekten zusammen. MathSEE fördert neue, interdisziplinäre mathematische Forschung am KIT und stärkt seine Sichtbarkeit innerhalb und außerhalb des KIT.

*Tanja Hagedorn<sup>1</sup>, Martin Frank*

Am 23. Oktober 2018 wurde mit einem Empfang im Atrium des Kollegengebäudes Mathematik die Gründung des neuen KIT-Zentrums Mathematics in Sciences, Engineering, and Economics (MathSEE) gefeiert (s. Foto), das bereits am 1. Oktober 2018 seine Arbeit aufnahm. Der Sonderforschungsbereich 1173 "Wellenphänomene: Analysis und Numerik" und andere bestehende Kooperationen bilden die Grundlage für die Entstehung von MathSEE.

Als international einzigartiger Rahmen bündelt MathSEE die bestehende und zukünftige interdisziplinäre mathematische Forschung am Karlsruher Institut für Technologie (KIT) zwischen Theorie und Anwendung sowie Universitäts- und Großforschung. Wissenschaftlerinnen und Wissenschaftler aus Mathematik sowie Natur-, Ingenieur- und Wirtschaftswissenschaften ab Masterabschluss arbeiten in MathSEE an interdisziplinären Forschungsfragen zusammen.

Ziel von MathSEE ist es, neue Forschungskooperationen zu generieren und vorhandene Forschung nach innen und außen sichtbar zu machen. MathSEE bietet dazu gezielte wissenschaftliche Veranstaltungen und Anschubförderung für Forschungsprojekte an. Das Zentrum hat inzwischen fast 90 Mitglieder aus 34 Instituten des KIT.

MathSEE ist in Bereiche mit unterschiedlichen methodischen Ansätzen strukturiert, zunächst in der folgenden Zuordnung:

- mathematische Modellbildung, Differentialgleichungen, Numerik, Simulation
- inverse Probleme, Optimierung
- mathematische Strukturen: Formen, Geometrie, Zahlentheorie und Algebra
- stochastische Modellbildung, statistische Datenanalyse und Vorhersage

Außerdem entsteht mit der zentrumsweiten Graduiertenschule MathSEED ein struktureller Rahmen für das interdisziplinäre Veranstaltungsangebot für die promovierenden Mitglieder. Vorhandene Lehrangebote, die für Masterstudierende und Promovierende anderer Fächer geöffnet werden können, werden identifiziert und Bedarfe nach zusätzlichen Qualifizierungsmöglichkeiten erhoben. Dies geschieht in Zusammenarbeit mit den promovierenden Mitgliedern und beteiligten Promotionsprogrammen. Alle Angebote von MathSEED sollen Promovierende zur Kooperation mit den jeweils anderen Disziplinen befähigen sowie Masterstudierende für die interdisziplinäre Arbeit begeistern.

<sup>1</sup> Geschäftsführerin, c/o Institut für Angewandte und Numerische Mathematik



In der ersten Mitgliederversammlung wurde Prof. Dr. Martin Frank, Direktor des SCC, zum Wissenschaftlichen Direktor von MathSEE gewählt (s. Foto). Er ist überzeugt, dass MathSEE mit seiner interdisziplinären Aufstellung einen weitreichenden Beitrag zur mathematischen Forschung am KIT und darüber hinaus leisten kann: „Die Mathematik als Schlüsseltechnologie der Digitalisierung hat einiges zu den Herausforderungen unserer Zeit beizutragen. Dazu zählen neben computergestützten Wissenschaften auch datengetriebene Methoden wie Maschinelles Lernen, die im Wesentlichen auf Statistik und Optimierung beruhen. Zudem erhoffen wir uns durch interessante Fragestellungen aus den Anwendungen Impulse für die Entwicklung neuer mathematischer Methoden.“

Die konkrete Arbeit in MathSEE ist bereits im vollen Gange. Das Kick-off im Methodenbereich Mathematische Strukturen hat stattgefunden, und die ersten weiterführenden Gespräche im Hinblick auf mögliche Kooperationen wurden vereinbart. Die Geschäftsstelle vermittelt bereits Kontakte. Weiterhin werden viele neue Mitglieder aufgenommen. Im Januar treffen sich die weiteren Methodenbereiche zum ersten Mal, und für 2020 ist eine große internationale Konferenz geplant.



■ Mehr Informationen unter: [www.mathsee.kit.edu](http://www.mathsee.kit.edu)

#### **KIT Center MathSEE “Mathematics in Sciences, Engineering, and Economics” has started its Work**

The recently founded KIT Center Mathematics in Sciences, Engineering, and Economics (MathSEE) pools the interdisciplinary mathematical research at KIT. It is set between theory and application as well as between university and large-scale research. Researchers from mathematics, natural sciences, engineering and economics work together on interdisciplinary projects. MathSEE fosters new interdisciplinary mathematical research at KIT and strengthens its visibility within KIT and beyond. To this end, MathSEE offers targeted scientific events and seed funding for research projects in an early phase. The graduate school MathSEED supports doctoral researchers in facing the challenges of interdisciplinary research. It opens and adapts existing courses and offers additional formats targeting interdisciplinary questions.

■ For more information, please visit the MathSEE website: [www.mathsee.kit.edu](http://www.mathsee.kit.edu)

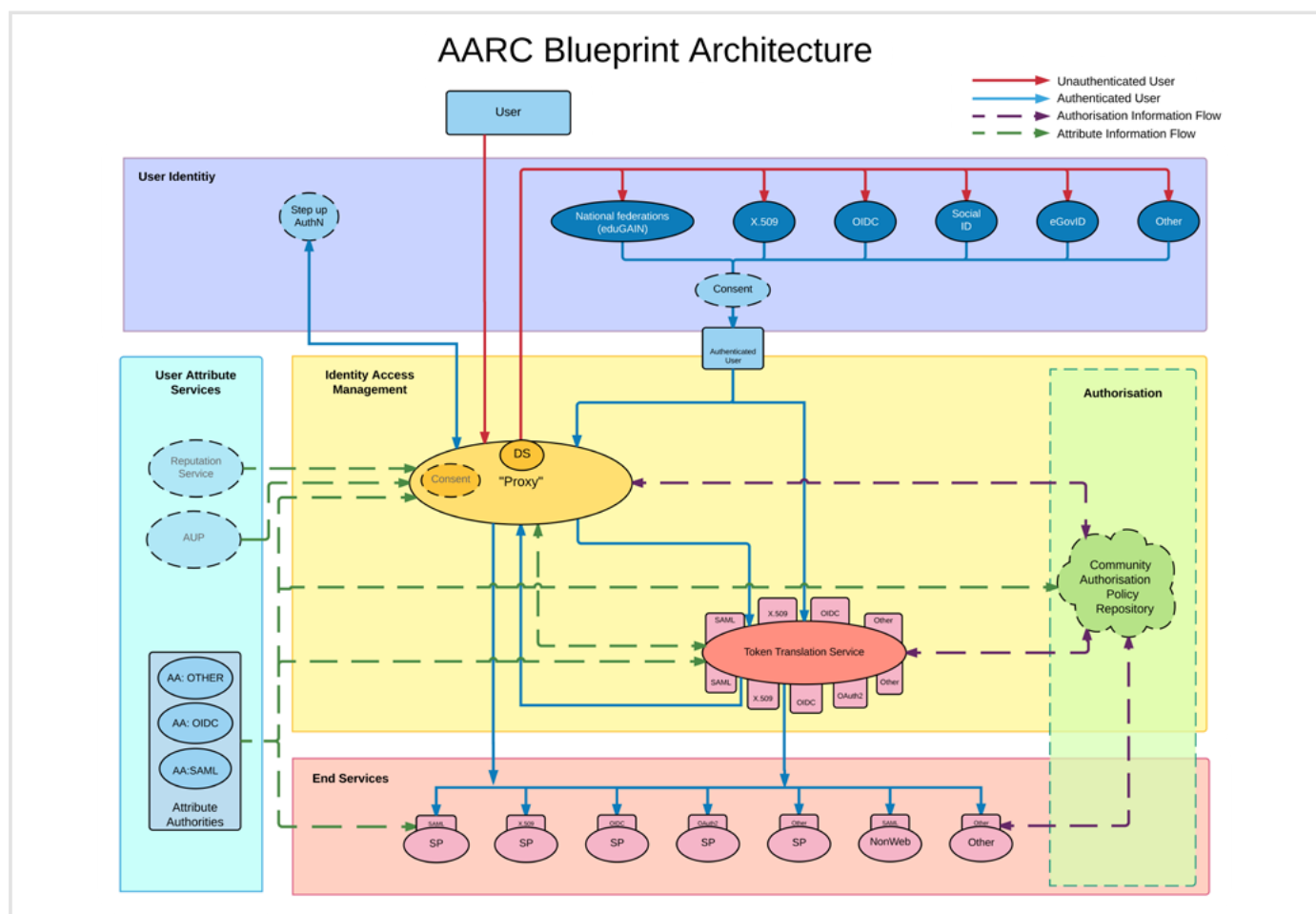
## Aufbau einer AAI für die Helmholtz Data Federation

Eine gut funktionierende Authentifizierungs- und Autorisierungsinfrastruktur, wie beispielsweise das föderierte Identitätsmanagementsystem bwIDM, ist die Grundlage für landesweit gemeinsam erbrachte und genutzte IT-Dienste. Das SCC hat sich in der Entwicklung und dem Betrieb solcher föderierten IDM-Systeme nun auch auf nationaler Ebene etabliert und spielt in unterschiedlichen Projekten eine wichtige Rolle bei der Weiterentwicklung von IDM-Lösungen. Die Helmholtz Data Federation (HDF) stellt neue Herausforderungen an Technologie und Kompatibilität im Bereich AAI.

Marcus Hardt, Michael Simon

Nach zweijähriger Projektlaufzeit startete im Juni 2013 der Produktivbetrieb von bwIDM, das föderierte Identitätsmanagementsystem für die Authentifizierungs- und Autorisierungsinfrastruktur (AAI) an wissenschaftlichen Einrichtungen in Baden-Württemberg. Auch heute, fünf Jahre später, ist die Wichtigkeit des bwIDM-Systems unbestritten, ermöglicht es doch den Mitgliedern der ca. 50 teilnehmenden Hochschulen und Universitäten den Zugang zu den in der Zwischenzeit aufgebauten Landesdiensten. Diese reichen von den sechs im Land betriebenen HPC-Clustern über bwSync&Share bis zu den landesweiten Archiv- und Storage-Diensten bwDataArchiv und SDS@HD.

Jedem, der sich mit dem rund um das bwIDM entstandenen Ökosystem beschäftigt, wird schnell klar, dass eine gut funktionierende AAI die Grundlage für die in Baden-Württemberg gemeinsam erbrachten und genutzten IT-Dienste ist. Das im SCC vorhandene Wissen über Entwicklung und Betrieb eines solchen föderierten IDM-Systems ist auch auf nationaler Ebene anerkannt und nachgefragt. Darüber hinaus nimmt das SCC an den EU-Projekten Authentication and Authorisation for Research and Collaboration (AARC), Indigo DataCloud und EOSC-hub sowie der Helmholtz Data Federation (HDF) eine wichtige Rolle bei der Weiterentwicklung von IDM-Lösungen ein.





Im Rahmen der Helmholtz Data Federation (HDF) erweitert das SCC nicht nur seinen Einzugsbereich über Baden-Württemberg hinaus, sondern setzt auch auf neue Technologien. Dabei bleibt jedoch stets die Kompatibilität mit den vorhandenen Infrastrukturen im Blick, so dass harte Brüche und Insellösungen vermieden werden.

Die HDF [1] ist der Zusammenschluss der Helmholtz Zentren AWI, DESY, DKFZ, GSI, FZJ und KIT, mit dem Ziel, eine föderierte Dateninfrastruktur für die Forschungsbereiche Energie, Erde und Umwelt, Gesundheit, Schlüsseltechnologien und Materie aufzubauen. In Arbeitspaket 1 ist das KIT zuständig für den Aufbau der AAI, der im Folgenden näher beschrieben wird.

### HDF-Architektur

Die AARC Blueprint Architectures (BPA) [2] sind ein wichtiges Ergebnis aus dem EU-Projekt AARC [3]. Sie sind aus der Analyse einer Reihe von verteilten AAI-Architekturen entstanden und beschreiben die AAI-Landschaft, in der sich AAI-Architekturen verorten lassen, so dass Softwarearchitekten und technische Entscheidungsträger bewährte Komponenten kombinieren können, um maßgeschneiderte Lösungen für ihre Anforderungen zu entwickeln. Ein wesentlicher Bestandteil der AARC Blueprint Architectures ist der SP-IdP Proxy, der zwischen den Diensten (Service Provider, SP) und den Home-IdPs liegt. Dieser Proxy ermöglicht es, einer Reihe von Anforderungen gerecht zu werden, die bislang nur schwierig oder gar nicht umgesetzt werden konnten. Es gibt bereits mehrere internationale Initiativen, die Bedarf an einem solchen SP-IdP Proxy haben. Daher existiert eine Reihe von SP-IdP-Proxy-Implementierungen. Für den Aufbau der HDF hat sich das Projektteam für das Produkt Unity entschieden. Unity wurde ursprünglich als Teil von UNICORE gestartet, dann von den EU-Projekten EUDAT/EUDAT-2020 zum SP-IdP-Proxy weiterentwickelt. Das Forschungszentrum Jülich (FZJ) betreibt für diese Projekte eine Reihe von Unity-Instanzen, auch die für die HDF-AAI. Mit Unity lassen sich eine Reihe von Anforderungen erfüllen, die für die HDF relevant sind. Dies sind im Einzelnen:

#### – Delegierte Gruppenverwaltung:

Die Verwaltung von Berechtigungsgruppen kann nicht von der Heimateinrichtung durchgeführt werden, weil diese in der Regel nicht die Möglichkeit hat, die Mitgliedschaft in einer Gruppe von internationalen Wissenschaftlern zu bestätigen. Stattdessen muss dies von der jeweiligen, sogenannten Virtuellen Organisation (VO) selbst getan werden. Das Konzept dieser VO sieht einen "Principal Investigator" vor, der einerseits die Mitgliedschaft in seiner VO organisiert, andererseits aber auch zuständig für die Organisation seiner VO-Ressourcen (Storage Quotas, Computing Time, ...) ist. Die Gruppenmitgliedschaft wird beim Login am Dienst mitgeliefert, so dass jeder Dienst selbst feststellen kann, zu welcher Gruppe ein Nutzer gehört und wie viele Ressourcen

noch für diese zur Verfügung stehen. Unity unterstützt diese Funktionalität bereits. Ein nutzerfreundliches Interface ist in Arbeit.

#### – Authentifizierung mit dem Home-IdP:

Für den Austausch von Authentifizierungs- und Autorisierungsdaten kommt der Standard SAML2 zum Einsatz. Das erlaubt die Authentifizierung über die DFN-AAI (Deutsches Forschungsnetz e.V.). Darüber hinaus werden auch Nutzer von einem Identitätsprovider (IdP) in EduGAIN [4] akzeptiert. Das ermöglicht es, Nutzer aus über 3000 IdPs weltweit an HDF-Diensten zu authentifizieren. Um selbst Nutzern den Zugang zu HDF-Diensten zu ermöglichen, deren Heimateinrichtung noch keinen IdP betreibt, sind auch ORCID und github (bald auch Google) als Identitätsprovider an Unity zugelassen. Hiermit geht dann das Problem einher, dass die Nutzer recht ungenau bekannt sind. Man spricht vom "Assurance Level", der hoch oder niedrig ist. Der Ansatz in HDF ist, die Assurance eines Nutzers mit dem "REFEDS Assurance Framework" [5] zu beschreiben. Diese Beschreibung ermöglicht es, die Assurance der Nutzer nach unterschiedlichen Gesichtspunkten (Komponenten) zu bewerten. Zusätzlich können gewisse Gruppen von Komponenten zu Profilen zusammengefasst werden. In HDF orientiert man sich an der DFN-AAI (vormals DFN-AAI-Advanced). Nutzer, die von einem DFN-AAI IdP kommen, erhalten so ein anderes Assurance-Profil als Nutzer aus dem Ausland oder solche, die Social Media verwenden. Die Dienste können dann selbst entscheiden, welche Nutzer sie akzeptieren. Um auch Nutzern mit einer niedrigen Assurance den Zugang zu Diensten zu gewährleisten, ist vorgesehen, dem VO-Manager die Identitätsprüfung zu übertragen. Details hierzu werden in den HDF-Policies [1] geregelt.

#### – Attributfreigabe:

Eines der häufigsten Probleme in föderierten Systemen ist die Attributfreigabe vom Identitätsprovider an die Dienste. Aus vorgelagerten Datenschutzgründen werden häufig überhaupt keine Attribute ausgeliefert. Die Verarbeitung und Weitergabe von personenbezogenen Daten wird von einer Reihe von rechtlichen Rahmenbedingungen, insbesondere die der Datenschutzgrundverordnung (DSGVO), geregelt. Erfreulicherweise gehen Datenschutzexperten (z.B. Andrew Cormack) davon aus, dass föderiertes Zugriffs- und Identitätsmanagement im wissenschaftlichen Bereich vom "Regulator" als Technologie zur Verbesserung der Privatsphäre anerkannt ist [7]. So ist es möglich, unter bestimmten Voraussetzungen und in eingeschränktem Umfang, persönliche Daten weiterzugeben. HDF folgt hier den AARC-Richtlinien zur DSGVO, die im Rahmen von AARC als "best practices" empfohlen werden [8]. Diese beinhalten zusätzlich Security und Community Management Policies und sind für HDF hier [6] zusammengefasst. Dies ermöglicht es Unity, einen definierten Satz an Attributen an die angeschlossenen Dienste weiterzugeben.

– Übersetzung zwischen unterschiedlichen Protokollen:

In der aktuellen föderierten Landschaft sind die meisten Nutzer in der Lage, sich über SAML2 zu authentifizieren. Viele Dienste, insbesondere sogenannte "non-web services", nutzen allerdings häufig andere Protokolle bzw. Authentifizierungsverfahren. SP-IdP-Proxies spielen hier eine entscheidende Rolle, weil sie einen bekannten Nutzer nun über andere Protokolle an die eigentlichen Dienste weiterreichen können. HDF setzt auf OpenID Connect [9], das auch von der DFN-AAI empfohlen wird. Auch SAML wird unterstützt, wenngleich derzeit keine Dienste darüber angeschlossen sind.

### HDF-Dienste

Zur Entwicklung der HDF-AAI wurde eine Test-Instanz von Unity aufgestellt, in die eine Reihe unterschiedliche Dienste integriert wurden. Diese sind im Einzelnen:

- Icinga Monitoring: Die HDF-Infrastruktur wird mit einer Icinga-Instanz, die die GSI betreibt, überwacht. Die Einbindung in die HDF-AAI erfolgt über das Apache-Modul "mod\_auth\_openidc". Es werden zwei VOs unterstützt.

- DCache WebDAV: Ein OIDC-authentisierter WebDAV-Zugang zu dCache wird vom DESY mit dem Prometheus-Dienst betrieben. Die Integration erfolgt dCache-spezifisch.
- OpenStack Test Service: Das SCC betreibt seit geraumer Zeit eine kleine OIDC-fähige Test-Instanz von OpenStack.
- WaTTS: Der Token Translation Service [10] wird ebenfalls am SCC betrieben. Er ermöglicht das Erzeugen von kurzlebigen, IGTF-kompatiblen (IOTA) Nutzerzertifikaten im X.509-Format. Das ermöglicht die Anmeldung an Grid-Diensten.
- FEUDAL: Das FEderated User Deployment portAL [11] wird aktuell am SCC entwickelt. Das Web-Portal verwendet Nutzer-Attribute sowie einen SSH-Key, um bei angeschlossenen Diensten Accounts für Nutzer zu erzeugen. Am KIT werden so Accounts in der LDAP-Facade von bwIDM erzeugt. Dies ermöglicht eine nahtlose Integration aller Systeme, die selbst keine föderierten Identitäten unterstützen, z.B. via ssh bei Unix-Systemen.

Eine aktuelle Liste der verfügbaren Dienste ist auf [12] verlinkt. Das SCC freut sich über die gute Zusammenarbeit unter den Zentren und den raschen Fortschritt der Aufbauarbeiten.

### Weblinks

- [1] [helmholtz.de/hdf](http://helmholtz.de/hdf)
- [2] [aarc-project.eu/architecture](http://aarc-project.eu/architecture)
- [3] [aarc-project.eu](http://aarc-project.eu)
- [4] [edugain.org](http://edugain.org)
- [5] [wiki.refeds.org/display/ASS/REFEDS+Assurance+Framework+ver+1.0](http://wiki.refeds.org/display/ASS/REFEDS+Assurance+Framework+ver+1.0)
- [7] [community.jisc.ac.uk/blogs/regulatory-developments/article/federated-access-management-and-gdpr](http://community.jisc.ac.uk/blogs/regulatory-developments/article/federated-access-management-and-gdpr)
- [8] [aarc-project.eu/guidelines/#policy](http://aarc-project.eu/guidelines/#policy)
- [9] [openid.net/connect](http://openid.net/connect)
- [10] [wiki.geant.org/display/AARC/WaTTS+RCauth+plugin++Obtaining+IOTA+proxy+certificates+with+OIDC](http://wiki.geant.org/display/AARC/WaTTS+RCauth+plugin++Obtaining+IOTA+proxy+certificates+with+OIDC)
- [11] [wiki.geant.org/display/AARC/FEderated+User+Deployment+PortAL+---+FEUDAL](http://wiki.geant.org/display/AARC/FEderated+User+Deployment+PortAL+---+FEUDAL)
- [12] [unity.helmholtz-data-federation.de/unitygw/VAADIN/files/connected-services.html](http://unity.helmholtz-data-federation.de/unitygw/VAADIN/files/connected-services.html)

### Establishment of an AAI for the Helmholtz Data Federation

A well-functioning system for person authentication and authorization such as bwIDM, the identity management system deployed in Baden-Württemberg, is the foundation for secure and nationwide use of commonly provided IT services. Development and operation of federated identity management systems at SCC already plays an important role in multiple international projects. Recently this activity was introduced at the national level and will support the upcoming federated data infrastructure of the Helmholtz Association. Technically this development challenges current implementations and compatibility with existing identity management solutions.

# Rechenzeitanträge für den ForHLR am SCC

Der Forschungshochleistungsrechner (ForHLR) des KIT steht Wissenschaftlerinnen und Wissenschaftlern aus ganz Deutschland für die Bearbeitung komplexer Anwendungsprobleme zur Verfügung. Bewerbungen für Rechenzeitprojekte werden über das Online-Formular eingereicht. Unmittelbar danach erhalten Sie Zugang zum System für erste Tests.

*Hartmut Häfner*

Für jedes begutachtete Rechenzeitprojekt ist eine Projektbeschreibung von 3-5 Seiten nötig. Die erste Seite beschreibt den wissenschaftlichen Hintergrund, außer das Projekt wurde erfolgreich wissenschaftlich begutachtet (z.B. DFG, BMBF, EU). Dann sind eine kurze Zusammenfassung und ein Verweis mit Förderkennzeichen und Link zur Projektseite ausreichend. Auf Seite 2 wird die zu verwendende Software beschrieben und insbesondere der Parallelisierungsgrad der verwendeten Applikation angeführt. Von Vorteil sind Skalierungsmessungen der Applikation, die den Speedup bei steigenden Prozessorzahlen zeigt. Diese können mit den Vorab-Konten vorgenommen werden. Außerdem sollte begründet werden, wofür gegebenenfalls spezielle Ressourcen wie GPU-Knoten oder ein Zugang zum Visualisierungslabor nötig sind. Auf Seite 3 werden die angegebenen CPU-Stunden und der benötigte permanente Plattenplatz begründet (siehe Beispiel rechts). Auf dem ForHLR I können maximal 15 Mio. CPU-Stunden und 4 TB Plattenplatz beantragt werden. Auf dem ForHLR II sind es 30 Mio. CPU-Stunden und 5 TB Festplattenspeicher.

Experten begutachten die eingereichten Projekte innerhalb von acht Wochen. Neben der wissenschaftlichen Relevanz ist ein wichtiges Kriterium, dass die Berechnungen eine große Anzahl von Prozessoren parallel benötigen. Das SCC unterstützt hierbei, sowohl während des Bewerbungsprozesses als auch über die Simulation Labs während der Projektlaufzeit. Die Dauer für Rechenzeitprojekte liegt zwischen ein und maximal zwei Jahren, mit der Möglichkeit der Verlängerung.

Nach der Bewilligung werden die Ressourcen umgehend gutgeschrieben. Die Projektmanager sind verpflichtet, am Ende des Bewilligungszeitraums einen 10 bis 15-seitigen Statusbericht zu erstellen.

Um kleinere Projekte ohne begutachteten Projektantrag zu ermöglichen, gewährt das SCC auf dem ForHLR I Test-Zugänge, die maximal ein Jahr gültig sind. Für diese Testprojekte werden maximal 500.000 CPU-Stunden zur Verfügung gestellt.

### Beispiel zur Ermittlung der benötigten CPU-Zeit und des permanenten Plattenplatzes:

Vereinfachend wird angenommen, dass parallele Programmläufe mit OpenFOAM auf 100 Knoten und 20 Cores pro Knoten – also auf 2000 Cores – durchgeführt werden, die jeweils 10 Stunden laufen. Während der Projektlaufzeit, die z.B. 18 Monate betragen soll, ist an jedem 2. Tag ein Programmlauf geplant. Es kämen ca. 275 Programmläufe zusammen, hier auf 300 aufgerundet, da auch Testläufe nötig sind oder Job-Abrüche vorkommen können. Die benötigte CPU-Zeit wäre: 300 (Läufe) x 2000 (Cores) x 10 (Stunden) = 6 Mio. CPU-Stunden.

Angenommen jeder Programmlauf erzeugt 10 MByte Daten, die über die Laufzeit des Projekts benötigt werden, kommt man so auf: 300 x 10 (MB) = 3 GB permanenten Plattenplatz.

### Die wichtigsten Links zu Rechenzeitanträgen:

<a href="http://www.scc.kit.edu/dienste/proposals.php">www.scc.kit.edu/dienste/proposals.php</a>	Beschreibung des Zugangsverfahrens auf Deutsch
<a href="http://wiki.scc.kit.edu/hpc/index.php/ForHLR_-_User_Access">wiki.scc.kit.edu/hpc/index.php/ForHLR_-_User_Access</a>	Beschreibung des Zugangsverfahrens auf Englisch
<a href="http://www.scc.kit.edu/forschung/4971.php">www.scc.kit.edu/forschung/4971.php</a>	Online Proposal Form
<a href="http://www.scc.kit.edu/forschung/10950.php">www.scc.kit.edu/forschung/10950.php</a>	Online Test Project Form
<a href="http://www.scc.kit.edu/servicedesk/formulare.php">www.scc.kit.edu/servicedesk/formulare.php</a>	ForHLR Zugangsformular

## Das SCC prägt die European Open Science Cloud

Das Mitwirken an internationalen Projekten steht am SCC auf Augenhöhe mit der Bereitstellung und Entwicklung von IT-Dienstleistungen für Forschung, Lehre und Administration. Bei Projekten, bei denen Planung, Konstruktion und Betrieb von komplexen Computer- und Datenumgebungen von zentraler Bedeutung sind, hat das SCC eine führende Rolle in Baden-Württemberg und im In- und Ausland inne. Neben den bewährten Rechenzentrumstätigkeiten ist das SCC vermehrt aktiv beim Aufbau und der Entwicklung von IT- und Service-Management Strukturen für die Integration von IT-Diensten aus unterschiedlichen Wissenschaftsdisziplinen und einem internationalen Umfeld. Diese Erfahrungen bringt das SCC besonders in den EU-Projekten EOSCpilot<sup>1</sup> und EOSC-hub<sup>2</sup> ein und prägt damit den Aufbau der integrierten europäischen Forschungsdateninfrastruktur.

*Eileen Kühn, Pavel Weber, Jos van Wezel*

Um die Vision der Europäischen Union, einen weltweiten Vorsprung im Forschungsdatenmanagement, zu erreichen, fördert die Europäische Kommission die Implementierung der European Open Science Cloud (EOSC). Die Kommission unterstützt somit einen grenzüberschreitenden, nahtlosen Zugang zu Datendiensten zur Speicherung, Verwaltung und Analyse von Forschungsdaten. Die EOSC ist sowohl ein Fortschrittskonzept als auch eine Implementierung. Sie wird Open Science ermöglichen und beschleunigt die digitale Transformation der Wissenschaft, weil mit den EOSC-Diensten Daten aus allen öffentlich finanzierten Forschungsarbeiten in Europa leicht zugänglich und wiederverwendbar sein werden. Die Finanzierung und Planung der EOSC stellt das EU-Rahmenprogramm für Forschung und Innovation (Horizon 2020) sicher.

### Das Gerüst der EOSC: EOSCpilot



Das im Januar 2017 gestartete EOSC-Pilotprojekt ist das erste einer Reihe von Projekten, mit denen eine allgemeine, offene, sichere und wirtschaftliche Forschungsdateninfrastruktur in Europa verwirklicht werden soll. Es bringt Forschungsgemeinschaften, -einrichtungen, -infrastrukturen einschließlich e-Infrastrukturen und die Forschungsförderer zusammen, um zur Entwicklung von EOSC beizutragen und diese zu lenken. Im Rahmen des Projektes EOSCpilot wurden die Erfahrungen aus früheren (EU-)Datenprojekten zusammengeführt, die initiale Lenkungsstruktur definiert, Richtlinien für den effektiven Betrieb und Zugang erarbeitet sowie eine Servicearchitektur entworfen. Das Projekt zeichnet sich besonders durch eine ehrgeizige Zusammenarbeit ausgewählter Forschungsgemeinschaften, den sogenannten Science Demonstrators, aus.

Die Science Demonstrators stellen einen Teil der zukünftigen EOSC-Communities dar. Sie demonstrieren den Nutzern die Möglichkeiten einer Dateninfrastruktur ohne Grenzen für die europäische Forschung, indem Ressourcen und Dienste internationaler Anbieter in bestehende Arbeitsabläufe der Forschungsdomänen

integriert werden. Gleichzeitig identifizieren die Science Demonstrators Verbesserungsmöglichkeiten, um zukünftige Projekte zu optimieren.

Ein weiteres Ergebnis des Projektes ist eine Reihe von Empfehlungen und begleitenden Maßnahmen zur Verwirklichung der EOSC. Sie enthalten Regeln für die Teilnahme an der EOSC (Rules of Participation<sup>3</sup>) und Empfehlungen zur Einbeziehung von Nutzern, zur Implementierung, zu FAIRen Daten, zu Datenschutz und Informationssicherheit, zur Beschaffung von Ressourcen<sup>4</sup> und mehr.

Das wichtigste Ergebnis ist zweifellos das Layout des initialen Lenkungsaufbaus, der die EOSC in die zweite Entwicklungsphase führen wird. Diese Phase hat zum Ziel, die Finanzierung, Nachhaltigkeit und ein langfristiges Management sicher zu stellen. Während des offiziellen EOSC-Launch<sup>5</sup> in Wien wurden am 23. November in einem ersten Schritt die Vertreter der wissenschaftlichen Organisationen zur Bildung des EOSC Executive Board offiziell benannt.

In EOSCpilot leitet das SCC das Arbeitspaket Services und trägt zur Daten-Interoperabilität im Arbeitspaket Interoperability sowie zur Aus- und Weiterbildung im Arbeitspaket Skills bei (Siehe dazu auch den Artikel zur GridKa School auf Seite 44).

### Aufbau der EOSC: EOSC-hub



Ein Jahr nach Start des Projektes EOSCpilot startete im Januar 2018 das Projekt EOSC-hub. Neben dem SCC sind in EOSC-hub erneut wichtige Infrastruktur-Anbieter aus dem EOSCpilot-Projekt sowie über 70 weitere Partner beteiligt. Der ehrgeizige Arbeitsplan zur Erstellung einer europaweiten föderierten Dateninfrastruktur, die

<sup>1</sup> [eosc-pilot.eu/](http://eosc-pilot.eu/)

<sup>2</sup> [eosc-hub.eu/](http://eosc-hub.eu/)

<sup>3</sup> [goo.gl/Vxdpck/](http://goo.gl/Vxdpck/)

<sup>4</sup> [eosc-pilot.eu/sites/default/files/prompting\\_an\\_eosc\\_in\\_practice\\_eosc\\_hleg\\_interim\\_report.pdf](http://eosc-pilot.eu/sites/default/files/prompting_an_eosc_in_practice_eosc_hleg_interim_report.pdf)

<sup>5</sup> [eosc-launch.eu/programme/](http://eosc-launch.eu/programme/)





auf den technischen Grundlagen von EGI.eu und den EUDAT-Projekten basiert, legt den Fokus auf die Integration von Daten-diensten und die Einrichtung einer Dienst-Management-Schicht auf Basis des Servicemanagement-Frameworks FitSM.

EOSC-hub arbeitet eng mit folgenden Projekten zusammen: einfraCentral<sup>6</sup>, bei dem die Organisation und Auffindbarkeit von wissenschaftlichen Datendiensten einen Schwerpunkt bildet, und OpenAIRE-advance<sup>7</sup>, das sich auf Dienste für Open Data und deren Förderung konzentriert. Wie auch EOSCpilot kooperiert EOSC-hub mit einer großen Anzahl von Wissenschaftsgemeinschaften. Diesmal werden die Communities jedoch aktiv dabei unterstützt, ihre Tools, Workflows, Dienste und mehr, in die Dateninfrastruktur einzubringen und sie mit der Föderationsschicht von EOSC zu verbinden.

Weniger als ein Jahr nach dem Start des Projektes gewinnt EO-SC-hub schnell an Dynamik. Die Präsentation des EOSC marketplace<sup>8</sup> während des erfolgreichen Launches von EOSC zeigt die konkrete Ausgestaltung der Forschungslandschaft der Verbund-forschungsdienste, zu der das SCC weiterhin beitragen wird.

In EOSC-Hub leitet das SCC das Arbeitspaket "Integration and maintenance of federation and collaborative services" und ist damit für die Integration von core services<sup>9</sup> verantwortlich. Es trägt zur Bereitstellung sowie zum Management von förderierten Diensten bei und ist an der Definition der Verwaltungsprozesse und Betriebsabläufe beteiligt. Zu den wichtigsten Aktivitäten, die vom SCC koordiniert werden, gehören die Integration von Authentifizierungs- und Autorisierungsdiensten, von Accounting- und Überwachungssystemen sowie von Diensten zur Unterstützung der Serviceplanung und -bereitstellung.

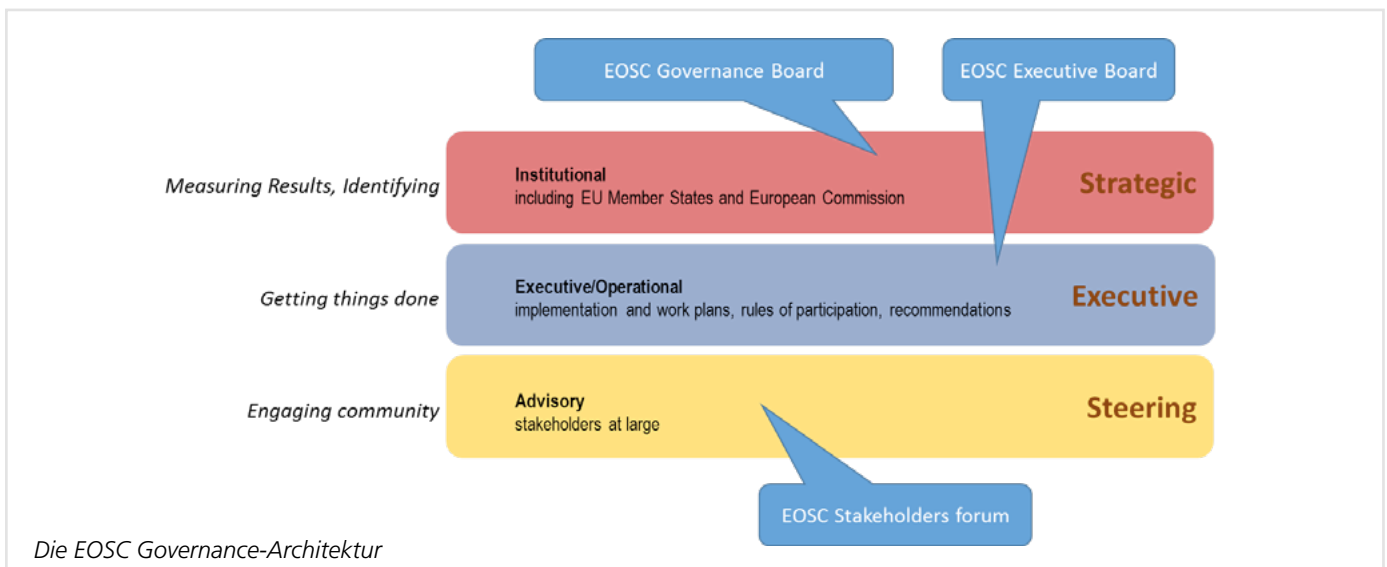
In der nächsten Ausgaben der SCC-News wird ausführlich über die Fortschritte und Ergebnisse des Projekts und die Arbeiten des SCC informieren.

<sup>6</sup> [einfracentral.eu/](http://einfracentral.eu/)

<sup>7</sup> [www.openaire.eu/advance/](http://www.openaire.eu/advance/)

<sup>8</sup> [marketplace.eosc-hub.eu/](http://marketplace.eosc-hub.eu/)

<sup>9</sup> [ec.europa.eu/research/openscience/pdf/swd\\_2018\\_83\\_f1\\_staff\\_working\\_paper\\_en.pdf](http://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf)



Die EOSC Governance-Architektur

### SCC shapes the European Open Science Cloud

Participation in international projects at SCC is on an equal level with the provision and development of IT services for research, teaching and administration. Where planning, design, construction and operation of complex computer and data environments are of central importance, SCC plays a leading role in in Baden-Württemberg, on the national and international level. In addition to the common data center activities SCC is increasingly implementing IT and service management structures along with the integration of services from different scientific disciplines. SCC contributes its experience especially in the EU projects EOSCpilot and EOCS-hub and through this, helps shaping the development of the future European research data infrastructure.

## MINTernationaler Erfolg des KIT

Das KIT konnte mit einem gemeinsamen Konzept von ZML, INTL, IAR und SCC einen Erfolg feiern. Für die Weiterentwicklung des Konzepts, bei dem es darum geht, MINT-Fächer zu Vorreitern der Internationalisierung zu machen, erhält das KIT 100.000 Euro vom Stifterverband, der Daimler und Benz Stiftung und dem Daimler Fonds.

*Ulrich Weiß*



Das Konzept des KIT war eines von drei Hochschulvorhaben, die sich unter den insgesamt 49 Teilnehmern des Wettbewerbs "MINTernational innovativ" durchsetzten. Die Hochschulen waren Ende Juni eingeladen, ihre Ideen für mehr Internationalität in den Fächern Mathematik, Informatik, Naturwissenschaften und Technik (MINT) einer Jury zu präsentieren. Zusammengesetzt aus Vertretern von Wissenschaftseinrichtungen und -förderern sowie der Wirtschaft, wählte diese die Konzepte mit dem höchsten Innovationsgrad aus. Mit den bereitgestellten Fördermitteln soll die jeweilige Projektidee gezielt weiterentwickelt, auf weitere Fächer ausgeweitet werden, um von anderen Hochschulen übernommen werden zu können.

Kai Rebensburg, der am KIT das Studierendenaustauschprogramm MINTernship betreut, und Andreas Sexauer vom Zentrum für Mediales Lernen (ZML) am KIT, entwickelten in Zusammenarbeit mit Ulrich Weiß vom SCC und Sebastian Stüker vom Institut für Anthropomatik und Robotik (IAR) ein Konzept, das es ausländischen Studierenden erleichtert, ihr Studium erfolgreich abzuschließen.

Hierzu wird mit den Mitteln des Innovationsfonds am KIT ein nachhaltiges Modell entwickelt, um anderen Hochschulen den Einsatz des Lecture Translators zum automatisierten Simultanübersetzen von Lehrveranstaltungen mit möglichst geringem Aufwand zu ermöglichen. Der Lecture Translator ist Resultat der Forschungen am Institut für Anthropomatik und Robotik (IAR) und übersetzt bereits seit 2012 in ausgestatteten Hörsälen des KIT Vorlesungen automatisiert in bis zu vier Sprachen.

Das SCC stellt zukünftig auf der virtuellen Serverumgebung einen performanten und skalierbaren Betrieb des Lecture Transla-

tors sicher. Der Dienst steht dabei auch zur asynchronen Verwendung zur Verfügung, bei der aufgezeichnete Veranstaltungen nachträglich transkribiert und übersetzt werden.

Am KIT wird die automatisierte Simultanübersetzung der Vorlesungen in andere Sprachen gemeinsam mit dem ZML zu einem regelmäßigen Serviceangebot weiterentwickelt. Dabei geht es nicht nur um den technischen Einsatz, sondern auch um Beratung von Lehrenden und Unterstützung bei der Verwendung des Systems. Mit dieser innovativen Eigenentwicklung reduziert das KIT die Sprachbarrieren für Studierende aus der ganzen Welt.

### "MINTernational" Success of KIT

KIT was able to celebrate a success with a joint concept of ZML, INTL, IAR and SCC. KIT will receive 100,000 euros from the Stifterverband, the Daimler und Benz Foundation and the Daimler Fund for the further development of the concept, which aims to make MINT subjects pioneers of internationalization.

The KIT concept was one of three university projects that prevailed among the 49 participants of the "MINTernational innovativ" competition. Within the framework of this concept, KIT is further developing the automated simultaneous translation of lectures into other languages together with ZML. This applies not only to the technical implementation, but also to advising teachers and supporting them in using the system. With this innovative in-house development, KIT is reducing the language barriers for students from all over the world.

## MINT-EC Akademie CAMMP week im September in Karlsruhe

Im September lockte die MINT-EC Akademie CAMMP week 24 Schüler/innen nach Karlsruhe, um sich eine Woche lang komplexen, relevanten und ungelösten Problemen aus Alltag, Industrie oder Forschung zu stellen. Organisiert und unterstützt wurde die Woche vom MINT-EC Verein, dem Projekt CAMMP, dem Projekt Simulierte Welten und dem Schülerlabor Mathematik des KIT.

*Kirsten Wohak, Maren Hattebuhr*

Die Problemstellungen, welche mit Hilfe der mathematischen Modellierung gelöst werden können, stammten von Mitarbeiterinnen und Mitarbeitern des SCC und des AIFB<sup>1</sup>. Die Schüler/innen beschäftigten sich mit folgenden Fragestellungen:

- Laufzeitvorhersage und Scheduling im High Performance Computing (Mehmet Soysal)
- Automatische Erkennung von Diagrammen in Kodizes (Dr. Danah Tonne)
- Optimales Entscheidungsverhalten für Interaktionen beim autonomen Fahren (Karl Kurzer)
- Optimale Kristallstruktur im Grundzustand (Momin Ahmad)

Nach der Begrüßung und einer ersten Erkundung der Stadt Karlsruhe am Sonntag, startete die Arbeit für die Schüler/innen am Montagmorgen mit einem Vortrag. Michael Weimer, Doktorand am SCC, führte die Teilnehmenden ausführlich in die mathematische Modellierung, am Beispiel der polaren Stratosphärenforschung, ein. Mit diesem Wissen stiegen die Schüler/innen in die Gruppenarbeit ein, um Lösungen für ihre komplexen Problemstellungen zu finden. Die wissenschaftliche Betreuung der Gruppen übernahmen Thomas Camminady, Jonas Kusch und Kirsten Wohak (SCC) sowie Jan Bohn (Mathematik). Die Aufgabe der Betreuungspersonen bestand darin, die Gruppen bei der mathematischen Umsetzung zu unterstützen. Den Lösungsweg für die Probleme kannten auch die Betreuungspersonen nicht, was die Arbeit für sie und die Schüler/innen gleichermaßen spannend machte. Zwangsläufig gehört zu dieser offenen Art der Forschung Frustration und auch das Hineinlaufen in Sackgassen. Dennoch blieben alle am Ball und arbeiteten eifrig weiter – teilweise sogar bis tief in die Nacht hinein.

Dienstagabends besuchten Problemsteller und weitere Interessierte die Schülergruppen in ihren Arbeitsräumen und diskutierten mit ihnen das bereits Geschaffte. Vielversprechende Ideen wurden in den folgenden Tagen umgesetzt, verworfen und neu entwickelt.

Während der Wanderung zum Durlacher Turmberg am Mittwochnachmittag genossen die Schüler/innen die Sonne und die Bewegung als gelungene Abwechslung zum Forschungsalltag. Inzwischen wurde parallel zu der Forschung daran gearbeitet, die bisherigen Ergebnisse in Form eines Berichts festzuhalten und strukturierte Folien für die Abschlusspräsentation zu erstellen. Am Donnerstagnachmittag hielten die einzelnen Gruppen jeweils einen Probevortrag.

Im Rahmen einer feierlichen Abschlussveranstaltung präsentierten die - zurecht stolzen - Schüler/innen ihre Ergebnisse im großen Hörsaal am SCC Campus Nord vor einem vollen Auditorium. Die Problemsteller/innen bedankten sich für die ausdauernde und erfolgreiche Arbeit der Schüler/innen. Sie waren allesamt beeindruckt, wie viel eine Gruppe von sechs motivierten jungen Menschen in nur fünf Tagen und der einen oder anderen Nacht erreichen kann. In den Vortragspausen, in denen es auch Kleinigkeiten zur Stärkung gab, diskutierten die Schüler/innen mit den Problemstellenden und weiteren Interessierten, bevor alle ihre Heimreise antraten.

Wieder ist eine erfolgreiche MINT-EC Akademie CAMMP week vorbei. Die nächste mathematische Modellierungswoche folgt im September 2019.

<sup>1</sup> Institut für Angewandte Informatik und Formale Beschreibungsverfahren



### MINT-EC Academy CAMMP week in Karlsruhe in September

In September 24 students participated in the MINT-EC Academy CAMMP week in Karlsruhe to face complex, relevant and unsolved problems for one week. The problems, which can be solved with the help of mathematical modelling, came from employees of SCC and AIFB. The solution for the problems was not known by the supervisors either, which made the work for them and the students equally exciting. Inevitably, this open type of research includes frustration and running into dead ends. Nevertheless, all the students stayed focused and continued to work eagerly. In a festive closing ceremony, the students - deservedly proud - presented their results to a full auditorium in the large lecture hall at SCC Campus North.

## 16. Internationale GridKa School – Computing and Science Fair

Vom 27. bis 31. August 2018 fand die bereits 16., vom SCC organisierte, internationale GridKa School statt. Mit bis zu 13 Stunden Plenarvorträgen und bis zu 85 Stunden Workshops bot die Sommerschule den Teilnehmern aus den unterschiedlichsten Fachrichtungen wieder ein sehr umfangreiches und spannendes Programm zu aktuellen Themen des Scientific Computing. Insgesamt nahmen in diesem Jahr mehr als 140 Wissenschaftlerinnen und Wissenschaftler aus 17 Nationen teil.

*Eileen Kühn, René Caspart*

Auch in diesem Jahr blickt das SCC auf eine weitere erfolgreiche GridKa School zurück, die unter dem Motto *Computing and Science Fair* organisiert wurde. Die 143 Teilnehmer aus 17 Nationen haben sich in der Woche vom 27. bis 31. August am KIT Campus Nord intensiv mit Themen des Scientific Computing beschäftigt. Insbesondere Trends aus dem High Performance, High Throughput und Cloud Computing sowie Themen der Datenanalyse standen im Fokus.

Um den Teilnehmerinnen und Teilnehmern einen guten Einstieg in das sehr breite Themenspektrum zu ermöglichen, sieht die GridKa School ein spezielles Format der Wissensvermittlung vor: Die Teilnehmer/innen erhalten vormittags eine Einführung in die diversen Themenbereiche der Sommerschule und haben nachmittags die Möglichkeit, neu erworbenes Wissen praktisch zu erproben oder ihr Wissen weiter zu vertiefen. Dazu können die Teilnehmenden täglich aus sechs parallelen Workshops wählen. Die Workshops decken ein breites Spektrum von Themen und Schwierigkeitsgraden ab.

In diesem Jahr bot die GridKa School in Zusammenarbeit mit den Projekten Helix Nebula – The Science Cloud<sup>1</sup> sowie Large Scale Data Management and Analysis<sup>2</sup> zwei dedizierte Tage, die sich aktuellen Trends widmeten: dem Cloud Computing sowie der großskaligen und effizienten Analyse von Daten. Im Rahmen der Plenarvorträge wurden mit Themen wie beispielsweise der Sicherheit im Bereich des Fog und Edge Computings, der Virtualisierung von Forschungsumgebungen, der Visualisierung von großen Datenmengen sowie den Aufgaben der Research Data Alliance unterschiedliche Kompetenzfelder der Teilnehmer/innen angesprochen. Auch im Rahmen der Workshops konnten sie tiefer in Spezialgebiete einsteigen. Dies waren unter anderem die Reproduzierbarkeit und das Management von Workflows, die Arbeit mit Container-Technologien, die Einführung in unterschiedliche Programmiersprachen und -konzepte sowie Machine Learning und Profiling für hochperformante Systeme. Das detaillierte Programm der GridKa School 2018 sowie die veröffentlichten Materialien der Vorträge und Workshops können unter der Webseite <http://gridka-school.scc.kit.edu/2018/> abgerufen werden.

Wie auch in den vorangegangenen Jahren war die als SCC-Kolloquium organisierte Evening Lecture wieder ein besonderes Highlight. Für den Abendvortrag mit dem Thema „Thirteen modern ways to fool the masses with performance results on parallel computers“ konnte dieses Jahr Dr. Georg Hager von der Universität Erlangen-Nürnberg gewonnen werden. Mit viel Witz und Humor stellte Dr. Hager den Teilnehmerinnen und Teilnehmern 13 Methoden vor, wie mit Hilfe grafischer Auswertungen Ergebnisse so präsentiert werden können, dass sie beim Betrachter den bestmöglichen (wenn auch nicht immer wissenschaftlich korrekten) Eindruck hinterlassen.

In Summe bot die GridKa School 2018 mit 13 Stunden Plenarvorträgen und 85 Stunden Workshops ein sehr umfangreiches Programm, das es den Teilnehmenden nicht immer einfach machte, ihr persönliches Schulungsprogramm zusammenzustellen. Daher ist es nicht verwunderlich, dass das Feedback nahelegt, dass „mindestens die doppelte Zeit notwendig [sei], an allen Workshops von Interesse teilzunehmen“. Für viele ist dies ein Grund, jährlich an der GridKa School teilzunehmen. Deshalb wurden 2018 erstmals diejenigen ausgezeichnet, die bereits sechs und sieben Jahre in Folge an der GridKa School teilgenommen haben. Dies ist nicht nur ein Zeichen für das Interesse an den Themen der Schule sowie die gleichbleibend hohe Qualität der Beiträge der Tutor/innen und Sprecher/innen, sondern auch eine besondere Auszeichnung für das gesamte Organisationsteam der GridKa School selbst. Speziell der über die Jahre hinweg fortwährende Einsatz des Teams sorgt für den großen Erfolg der GridKa School.

Das vorrangige Ziel der GridKa School ist es allerdings nicht, dass Personen möglichst häufig an der Sommerschule teilnehmen, sondern dass sie die für sie relevanten Kernkompetenzen erwerben und sich diese möglichst direkt in ihrem Arbeitsumfeld adaptieren lassen. Hierzu ist es notwendig, dass die Teilnehmenden basierend auf ihren aktuellen Wissensbedarfen und Zielen die bestmöglichen Workshops aus dem breiten Angebot der Sommerschule finden und wählen können. Um dies künftig besser zu ermöglichen, wurden spezielle Umfragen für Teilnehmer/innen und Tutor/innen erstellt, um die Anwendbarkeit und

<sup>1</sup> [www.hnscicloud.eu](http://www.hnscicloud.eu)

<sup>2</sup> [www.helmholtz-lsdma.de](http://www.helmholtz-lsdma.de)





Teilnehmerinnen und Teilnehmer der GridKa School 2018, Copyright Christoph Heidecker

Möglichkeiten des im Rahmen von EOSCpilot<sup>3</sup> entwickelten Kompetenzframeworks FAIR4S<sup>4</sup> (siehe Kasten) zu validieren. Ein Einsatz des Kompetenzframeworks könnte nicht nur die Planung des Schulungsprogramms verbessern, sondern dem Organisationsteam ein Tool an die Hand geben, das Programm gezielter auf aktuelle Bedarfe und Kompetenzfelder abzustimmen.

Neben den Plenarvorträgen und Workshops fördert das Rahmenprogramm der Schule den interdisziplinären Austausch und gibt Gelegenheit zum Erfahrungsaustausch zwischen Teilnehmer/innen, Sprecher/innen und Tutor/innen. Zu diesem Rahmenprogramm zählte nicht nur der traditionelle Flammkuchenabend mit Führung durch das GridKa Computing and Data Centre und

den Forschungshochleistungsrechner (ForHLR), sondern auch das GridKa School Dinner. Beide Veranstaltungen sowie die zahlreichen Kaffeepausen wurden intensiv für vertiefende Gespräche und ungezwungene Diskussionen genutzt.

Die nächste GridKa School wird vom 26. August bis 30. August 2019 am KIT Campus Nord stattfinden. Interessierte können noch bis zum 28. Februar 2019 Vorschläge für Vorträge und Workshops einreichen ([indico.scc.kit.edu/event/460/](https://indico.scc.kit.edu/event/460/)) Das Themenspektrum der GridKa School 2019 reicht von modernen Programmier- und Analysetechniken über Machine Learning und statistischen Methoden im Data Mining, Fehlertoleranz und Reliability bis hin zu Methoden für skalierbare und hochperformante Systeme.

### Das FAIR4S Kompetenzframework

Das *framework of FAIR data stewardship skills for science and scholarship* FAIR4S wird im Rahmen des Arbeitspakets Skills im Projekt EOSCpilot entwickelt. Es listet die Fähigkeiten und Kompetenzen auf, um mit Forschungsobjekten, wie Daten oder Software, FAIR (Findable, Accessible, Interoperable, Reusable) zu arbeiten. Das Framework soll unterschiedliche Zielgruppen ansprechen und möchte beispielsweise bei der Spezifikation von Kompetenzen zur effizienten Nutzung von Services, der Identifikation relevanter Trainingsmaterialien und -inhalte sowie der Karriereplanung unterstützen. Um diese Ziele zu erreichen, unterscheidet das Framework unterschiedliche Dimensionen, um die Expertise, die Zielgruppe sowie die Verantwortlichkeit zu definieren und eine Bewertung zu ermöglichen.

Das Kompetenzframework ist aktuell in Version 1.0 verfügbar<sup>5</sup>. Kommentare und Anmerkungen zur Entwicklung des Frameworks sind ausdrücklich erwünscht. Es wurde evaluiert, ob FAIR4S auch für die GridKa School einsetzbar ist. Die Ergebnisse dieser Auswertung sowie weiterer Workshops und deren Bewertung für die Praktikabilität des Kompetenzframeworks wurden im Rahmen des Reports "D7.4: Report on Training Workshops"<sup>6</sup> auf der EOSCpilot-Webseite veröffentlicht.

### 16th GridKa School – Computing and Science Fair

The 16th international GridKa School, hosted by the Steinbuch Centre for Computing, took place from 27th to 31st of August. More than 140 participants from 17 countries with a broad range of expertise in scientific computing attended this year's summer school at KIT. This year's programme covered recent topics in scientific computing ranging from cloud computing to high throughput and high performance computing to data analysis.

<sup>3</sup> eoscpilot.eu

<sup>4</sup> eoscpilot.eu/sites/default/files/fair4s\_eoscpilot\_skills\_framework.pdf

<sup>5</sup> eoscpilot.eu/themes/wp7-skills

<sup>6</sup> eoscpilot.eu/content/d74-report-training-workshops





Reges Interesse der Teilnehmenden bei 13 Stunden Plenarvorträgen.\*



Auch die Führungen durch das GridKa und den ForHLR II sorgten für Begeisterung bei den Teilnehmenden.\*



Großartige Mitarbeit der Teilnehmerinnen und Teilnehmer bei Workshops von insgesamt 85 Stunden.\*







Besucher im Rechencluster.\*

## 16. Internationale GridKa School – Impressionen



Ungezwungene Diskussionen und Erfahrungsaustausch zwischen Teilnehmern und Tutoren während des umfangreichen Rahmenprogramms.\*



Auszeichnung von Michael Bontenackels und Aleksander Paravac für die Teilnahme seit sieben bzw. sechs Jahren in Folge.\*

\*Bilder Copyright Christoph Heidecker

## Erster Doctoral Science Slam am SCC

Am 10. Oktober 2018 fand am SCC der erste Doctoral Science Slam statt. Eine Doktorandin und fünf Doktoranden aus den unterschiedlichen Forschungsgruppen des SCC traten an, um ihre aktuellen Forschungen und Ergebnisse in allgemein verständlicher, humorvoller und kompakter Art und Weise vorzustellen. Knapp 50 interessierte Kolleginnen und Kollegen des SCC besuchten diese unterhaltsame Wissenschaftsveranstaltung.

*Eileen Kühn, Marco Berghoff*

Falls Sie nicht wissen, was das „grüne Versace Kleid von Jennifer Lopez“, die „Teenage Mutant Nano Robots“, „Monsteralarm“ bei der Entstehung von Salpetersäure oder „Farts at Exascale“ mit den Forschungen am SCC zu tun haben, dann haben Sie offensichtlich eine großartige Auftaktveranstaltung des Doctoral Science Slam am 10. Oktober 2018 verpasst. Sechs Promovierende aus unterschiedlichen Forschungsgruppen des SCC – Germaine Götzelmann, Pratik Nayak, Mehmet Soysal, Oskar Taubert, Michael Weimer und Thorsten Zirwes – präsentierten ihre aktuellen Forschungen und Ergebnisse den Mitarbeiterinnen und Mitarbeitern des SCC.

Die Wikipedia definiert: „Ein Science-Slam ist ein wissenschaftliches Kurzvortragsturnier, bei dem Wissenschaftler ihre Forschungsthemen innerhalb einer vorgegebenen Zeit vor Publikum präsentieren“, und so stellten sich auch die Promovierenden des SCC dem Wettkampf und dem Publikum. Ihre zehnminütigen populärwissenschaftlichen Vorträge wurden durch das Publikum bewertet, wobei Verständlichkeit sowie Unterhaltungswert des Kurzvortrags im Vordergrund standen.

Für den Auftakt des ersten Doctoral Science Slam konnte die erfolgreiche Science Slammerin Dr. Anastasia August vom KIT-Institut für Angewandte Materialien gewonnen werden. Anastasia präsentierte ihren neuen Vortrag zum Thema „Wärmewiderstand überwinden“ (Bild oben links). Mit diesem Vortrag gewann sie im Anschluss die süddeutsche Science Slam Meisterschaft und wurde im deutschen Science Slam Finale in Wiesbaden eine der Vizemeister/innen. Die teilnehmenden Promovierenden des SCC hielten großartige Vorträge, so dass es für das Publikum entsprechend schwer war, eine Entscheidung zu treffen und den besten Slammer zu küren. Gewinner des ersten Doctoral Science Slam ist Michael Weimer (Bild Mitte) mit seiner Präsentation zur „Simulation von polaren Stratosphärenwolken (PSCs) mit Icon-ART“. Nach seinem Vortrag verstehen nun auch die Kolleginnen und Kollegen am SCC die Wichtigkeit seiner wissenschaftlichen Ergebnisse und warum man manchmal „froh [sein kann], ein Ozonloch zu sehen“.

Nicht nur für Michael, sondern für alle an der Veranstaltung Beteiligten, war der erste Doctoral Science Slam ein Sieg: Die Vortragenden konnten ihre Forschung vor einem breiten Publikum vorstellen und dabei insbesondere lernen, wie man seine Forschungsfragen und -antworten nicht nur für Experten der eigenen Domäne präsentiert. Alle Zuschauer aus Forschung, Betrieb und Service erhielten einen einzigartigen Einblick in die aktuellen wissenschaftlichen Arbeiten, für und vom SCC. Der Doctoral Science Slam hat damit sicherlich dazu beigetragen, das Verständnis zwischen den einzelnen Forschungsgruppen sowie dem Betrieb und Service des SCC zu erhöhen und damit die gruppenübergreifende Kommunikation und Zusammenarbeit, speziell auch zwischen den Promovierenden, zu fördern. Vielleicht traut sich ja in Zukunft ein SCC-Slammer bei einem öffentlichen Science Slam gegen Anastasia anzutreten?

Rückblickend war der Doctoral Science Slam ein großartiger Auftakt für eine Reihe spannender Veranstaltungen. Wir haben sehr viel von unseren Kollegen und Kolleginnen gelernt und freuen uns jetzt schon auf den nächsten Doctoral Science Slam mit weiteren interessanten Themen.

### Durchweg positives Feedback

Das Feedback des Publikums während und nach der Veranstaltung war durchweg positiv. Beispielsweise bedankte sich Veronika Leonhardt für diese Veranstaltung. Am SCC kümmert sie sich um die Finanzen und Administration unterschiedlicher Projekte. „Mit den durch die unterhaltsamen Vorträge gewonnenen Einblicken kann ich mich jetzt noch besser mit den Forschungsarbeiten des Instituts identifizieren.“

■ \*Bilder Copyright Marco Berghoff





Wie mit Metall-  
schaum – hier im  
Modell – Wärmewi-  
derstand überwin-  
den werden kann,  
erläutert Dr. Anastasia  
August (KIT/IAM).\*



Über „Teenage Mutant Nano Robots“ slamt Oskar Taubert.\*



Germaine Götzelmann:  
„Von Buchdruckern und Bildersuchen –  
Buchillustrationen vor 500 Jahren auf der  
Spur“.\*



Michael Weimer zum Thema „Simulation  
von polaren Stratosphärenwolken (PSCs)  
mit Icon-ART“.\*



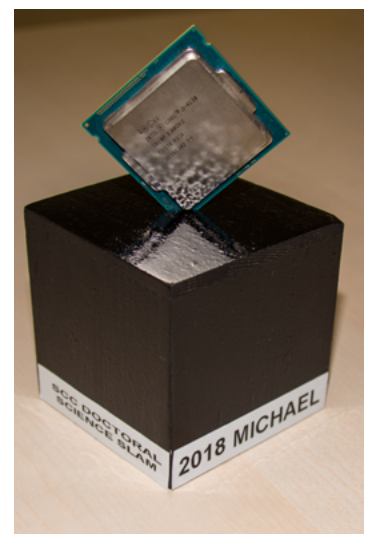
Thorsten Zirwes hat „brennende Fragen“  
... und Antworten. Sein Thema: „Burning  
Questions about Flames“.\*



Zum Glück geruchlos! Pratik Nayaks  
Thema „Studying farts at exascale:  
Domain decomposition methods“.\*



Ganz locker erklärt Mehmet Soysal wie  
„Bypassing the I/O Bottleneck“ funktioniert.\*



Wanderpokal des Doctoral  
Science Slam.\*

## Reinhard Strebler geht in den Ruhestand

Der Leiter der Abteilung Netze und Telekommunikation (NET), Reinhard Strebler, ist seit Ende 2018 im wohlverdienten Ruhestand. Nach einer Verlängerung seiner Dienstzeit von drei Monaten, blickt Reinhard nun auf runde 40 spannende und herausfordernde Jahre am KIT und der Universität Karlsruhe zurück. Die fachliche Leitung von NET übergab er vertrauensvoll schon in 2018 an seine Stellvertreterin Klara Mall, neuer Abteilungsleiter wird Klaus Scheibenberger, der auch weiterhin verantwortlich für die Abteilung Systeme und Server (SYS) ist.

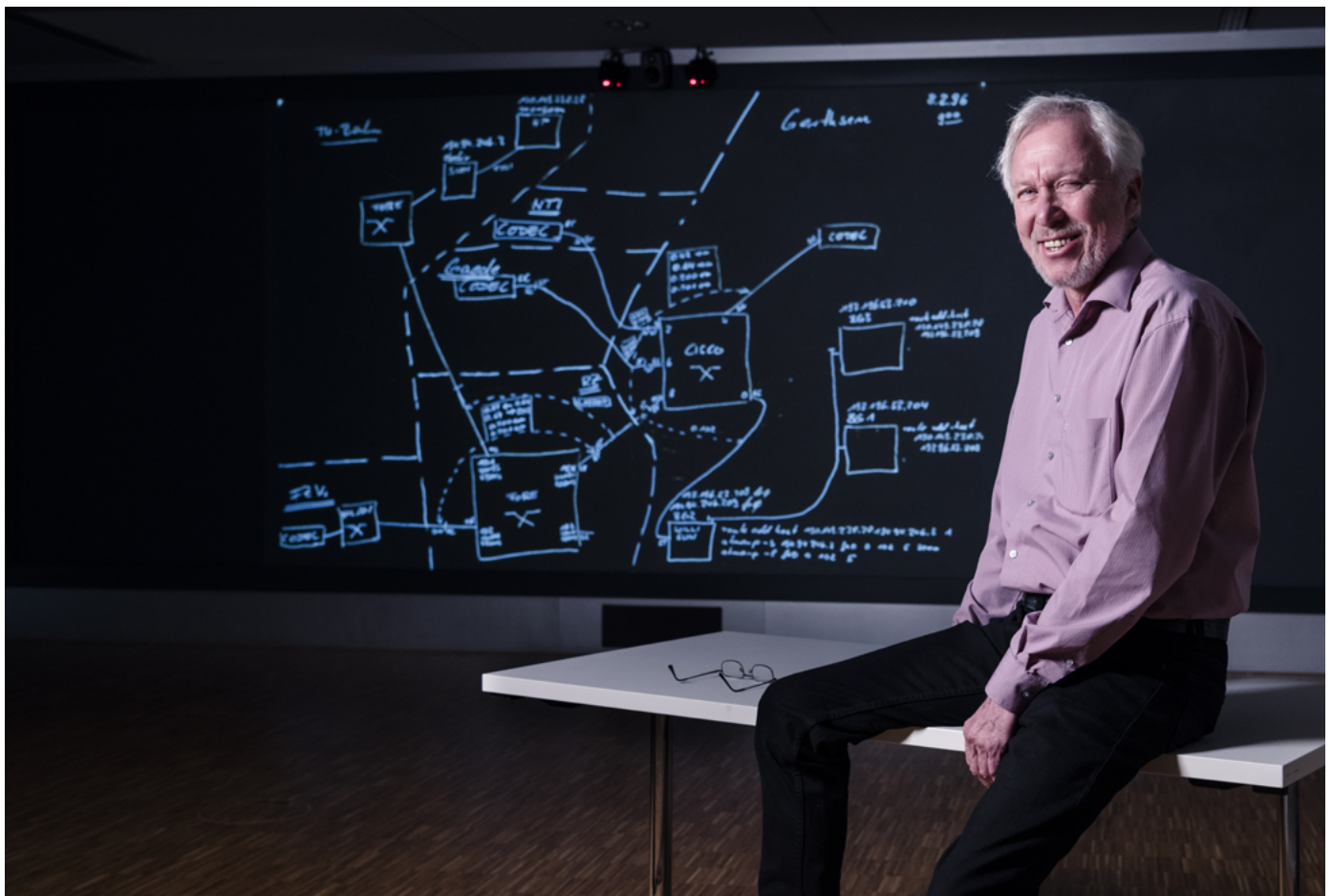
Bereits 1978 war Strebler, als frisch gebackener Elektrotechnikingenieur, am Rechenzentrum der Universität Karlsruhe

angestellt, in dem er schon zwei Jahre zuvor als studentische Hilfskraft mitwirkte. Dort, wo sie der Digitaltechnik überlegen waren, rechnete man damals mit Analog- und Hybridanlagen. Die Rechner-technik entwickelte sich erst Schritt für Schritt, und die Informatik war noch eine blutjunge Wissenschaft. Hier konnte sich der leidenschaftliche E-Techniker voll ausleben. 10 Jahre später wurde er Leiter der Abteilung Betrieb, Betriebssysteme und Technik am Uni-Rechenzentrum. Allmählich etablierte sich auch die digitale Netzwerktechnik in den Rechenzentren. Als sich 2008 die beiden RZs von Uni und Forschungszentrum Karlsruhe zum SCC zusammenschlossen und unter diesem gemeinsamen Dach neu strukturiert

wurden, übernahm Reinhard Strebler die Leitung der Abteilung Netze und Telekommunikation.

Der Netzwerkpionier Strebler hat die bahnbrechenden technischen Entwicklungen in der Computer- und Netzwerktechnik selbst miterlebt und berichtet ausführlich in dieser und der nächsten SCC-News über Erfahrungen und gemeinsame Erfolge mit seinem hervorragenden NET-Team.

Wir danken Reinhard für sein langjähriges Engagement und wünschen ihm alles Gute für den Ruhestand. (ag)



Reinhard Strebler, Netzwerkspezialist und Autor des Artikels auf Seite 4, zeigt das Arbeitspapier für die Konfiguration des ATM-Netzes anlässlich des Kolloquiums mit Bill Gates, 1996.



## Neues aus den SCC-Abteilungen

### Neue Abteilungsleiter – Philip Hoyer, ISL und Ulrich Weiß, AMA

Durch die Zusammenführung des SCC und der Dienstleistungsabteilung Allgemeine Services Verwaltungs-IT (ASERV-VIT) zu einer zentralen wissenschaftlichen Einrichtung entstanden neue Abteilungen. Wir begrüßen sehr herzlich Philipp Hoyer und Ulrich



Philip Hoyer

Weiß als neue Abteilungsleiter am SCC und wünschen ihnen viel Erfolg. Herr Hoyer ist verantwortlich für die Informationssysteme Studium und Lehre (ISL), Herr Weiß für die mit AMA abgekürzte Abteilung Anwendungen, Middleware und IT-Architektur.



Ulrich Weiss

## Auszubildender des SCC unter den Jahrgangsbesten

*Der Auszubildende Patrick Welter wurde bei der Jahresbestenehrung der IHK Karlsruhe für herausragende Leistungen in seiner Ausbildung als Fachinformatiker Systemintegration geehrt. Er absolvierte seine Ausbildung am SCC.*

Die IHK Karlsruhe (Industrie- und Handelskammer) führt jährlich eine Ehrung der besten Auszubildenden des Landkreises Karlsruhe durch. Die Feierstunde fand in der Gartenhalle Karlsruhe statt, bei der Ehrungen von rund 200 Auszubildenden in ca. 140 Ausbildungsberufen durchgeführt wurden.

Unter den Geehrten war auch SCC-Azubi Patrick Welter. Patrick absolvierte seine Ausbildung in der Abteilung Client-Management und Kommunikationsdienste (CMK), war aber auch in den Abteilungen ISM, NET und SYS des SCC tätig. Bei ISM unterstützte Patrick das ServiceDesk-Team, wo er mit seiner freundlichen und hilfsbereiten Art bei den Nutzern und im Team sehr beliebt war. Auch bei personellen Engpässen half er gerne aus. In der Netz-Abteilung lernte er verschiedene Netzwerkkomponenten und deren Konfiguration kennen. Bei SYS war er beim Aufbau und der Inbetriebnahme von Racks für Server und Storages beteiligt.

In seiner Ausbildungsabteilung CMK war er im Exchange-Team tätig, wo er von Michael Willhauk betreut wurde. Dort lernte er die Exchange-Organisation kennen, bearbeitete Nutzeranfragen zu Exchange und Outlook und erstellte Powershell-Skripte, um Arbeitsabläufe zu automatisieren.

In seiner Abschlussarbeit evaluierte Patrick die Überwachung von Exchange 2016 Server mittels SCOM 2016 (System Center Operations Manager) in einer Testumgebung. Die Arbeit umfasste die Installation und Konfiguration von SCOM sowie die Einstellung von Schwellwerten auf den zu überwachenden Systemen und die daraus resultierende Alarmierung von Fachpersonal. (ag)

### Weitere Informationen:

[www.karlsruhe.ihk.de/kommunikation/Pressemeldungen/ihk-karlsruhe-ehrt-auszubildende/4249580](http://www.karlsruhe.ihk.de/kommunikation/Pressemeldungen/ihk-karlsruhe-ehrt-auszubildende/4249580)

Berufsausbildung am KIT: [www.peba.kit.edu](http://www.peba.kit.edu)

## IMPRESSUM

### SCC news

Magazin des Steinbuch Centre for Computing

### Herausgeber

Präsident Professor Dr.-Ing. Holger Hanselka  
Karlsruher Institut für Technologie (KIT)  
Kaiserstraße 12  
76131 Karlsruhe

### Anschrift

Steinbuch Centre for Computing (SCC)  
Karlsruher Institut für Technologie (KIT)  
Redaktion SCC-News  
Zirkel 2  
76131 Karlsruhe  
oder:  
Hermann-von-Helmholtz-Platz 1  
76344 Eggenstein-Leopoldshafen  
Fax: +49 721 608-24972

### Redaktion

Achim Grindler (ag, verantwortlich),  
Karin Schäufele  
E-Mail: [redaktion@scs.kit.edu](mailto:redaktion@scs.kit.edu)

### Gestaltung, Satz und Layout

Hella Grolmus  
AServ – Crossmedia – Grafik  
Karlsruher Institut für Technologie (KIT)  
Hermann-von-Helmholtz-Platz 1  
76344 Eggenstein-Leopoldshafen

### Titelfoto

Amadeus Bramsiepe

### Fotos

Amadeus Bramsiepe, Marco Berghoff,  
Markus Breig, Maren Hattebuhr,  
Christoph Heidecker,

### Druck

Systemedia GmbH, 75449 Wurmberg

### Erscheinungstermin dieser Ausgabe

Februar 2019

[www.scc.kit.edu/publikationen/scc-news](http://www.scc.kit.edu/publikationen/scc-news)

Der Nachdruck und die elektronische Weiterverwendung sowie die Weitergabe von Texten und Bildern, auch von Teilen, sind nur mit ausdrücklicher Genehmigung der Redaktion gestattet.



Karlsruher Institut für Technologie (KIT)  
Steinbuch Centre for Computing (SCC)

ISSN: 1866-4954

[www.scc.kit.edu](http://www.scc.kit.edu)  
[contact@scc.kit.edu](mailto:contact@scc.kit.edu)