



Software Suite stellt opportunistische Ressourcen bereit
Software Suite provides opportunistic Resources

Workflows vom Fließband für die Materialforschung
Materials Science Workflows made easy

Research Software Engineering in der Aus- und Weiterbildung
Research Software Engineering in Education and Advanced Training

Liebe Leserinnen und Leser,

in den vergangenen Wochen war viel vom olympischen Sportsgeist und auch vom Motto der Spiele „Schneller, Höher, Stärker – Gemeinsam“ zu hören. Im übertragenen Sinn ist auch diese Ausgabe der SCC-News von einem solchen „Sportsgeist“ durchdrungen. „Gemeinsam innovativer, nachhaltiger und produktiver“ ist die inhaltliche Klammer, die sich über die hier vorgestellten Beiträge spannt. Spitzenplätze in puncto Energieeffizienz und Schnelligkeit hat der am 30.7. offiziell eingeweihte Hochleistungsrechner HoreKa bereits im Juni erreicht (S. 18). Damit das SAP-System des KIT auch in Zukunft Spitzenplätze erzielen kann, absolviert es derzeit ein Trainingslager (S. 4). Und um „gemeinsam“ die Präsenzlehre am KIT sicherer zu machen, wurde das elektronische Kontaktnachverfolgungssystem KONKIT (S. 20) entwickelt.

Das Research Software Engineering (RSE), das die Wissenschaft optimal in der Entwicklung von Forschungssoftware unterstützt, nimmt die Aspekte Innovation, Offenheit, Nachhaltigkeit und Produktivität besonders in den Fokus. Dafür braucht es einen fruchtbaren Austausch zwischen den Fachleuten, den die Veranstaltungsreihe SORSE auch in Pandemiezeiten sehr produktiv ermöglichte (S. 30). Eine lohnende Vorgehensweise ist es auch, Studierende rechtzeitig an die Methoden der Entwicklung nachhaltiger Forschungssoftware heranzuführen. Die praktische Umsetzung stellen wir Ihnen auf Seite 44 vor.

Mit Innovations- und Teamgeist hat eine kleine Entwickler-Gruppe eine Software Suite geschaffen, mit der Rechenressourcen aus Grid-, Cloud- und HPC-Infrastrukturen intelligent zusammengeführt und für Forschende transparent als virtuelles Cluster bereitgestellt werden (S. 8). Doch schnell und effizient rechnen ist in der Wissenschaft nur ein Erfolgsfaktor, auch deren Arbeitsabläufe sollten möglichst produktiv sein. Eine Projektgruppe des von der DFG geförderten Graduiertenkollegs GRK 2450 hat hierzu ein Open-Source-Werkzeug entwickelt, das Workflow-Anwendungen für die rechnergestützte Materialforschung automatisiert erstellen kann (S. 35).

Viel Freude beim Lesen

Martin Frank, Bernhard Neumair, Martin Nußbaumer, Achim Streit

Dear reader,

in the past weeks there has been a lot of talk about the Olympic spirit of sports and also about the motto “Faster, Higher, Stronger - Together”. In a figurative sense, this issue of SCC News is equally driven by such “sporting spirit”. The motto “Together more innovative, sustainable, and productive” frames the articles presented here. The high-performance computer HoreKa was officially inaugurated on July 30 and already achieved top rankings for its energy efficiency and speed in June (p. 18). To ensure that the KIT SAP system also continues achieving top rankings in the future, it is currently undergoing a training camp (p. 4). In order to make face-to-face teaching “together” at KIT more secure, the KONKIT electronic contact tracking system has been developed (p. 20). Research Software Engineering (RSE), which fosters the development of scientific research software, particularly focuses on innovation, openness, sustainability, and productivity. Even in times of pandemic, the Series of Online Research Software Events (SORSE) maintained the vital dialogue between experts (p. 30). A worthwhile approach is also to introduce students early to the methods of developing sustainable research software (p. 44).

With innovation and team spirit, a group of RSEs has created a software suite that smartly combines computing resources from grid, cloud and HPC infrastructures to transparently provide them to researchers as a virtual cluster (p. 8). Yet fast and efficient computing is only one success factor in science; its workflows should also be as productive as possible. To this end, a project group of the DFG-funded Research Training Group GRK 2450 has developed an open-source tool that automatically creates workflow applications for computational materials research (p. 35).

Enjoy reading

Martin Frank, Bernhard Neumair, Martin Nußbaumer, Achim Streit



DIENTE UND INNOVATION

- 04 Schritt für Schritt hin zu SAP S/4HANA
- 06 Einführung der elektronischen Ausgangsrechnung XRechnung
- 07 Aktionstag zum European Cyber Security Month 2020 am KIT
- 08 Bis zum GridKa und noch viel weiter – Opportunistische Ressourcen mit COBaID/TARDIS
- 12 Innovative Multifaktor-Authentifizierung für sicheren HPC-Zugang
- 15 Performanzverbesserung des GridKa-Magnetbandspeichersystems
- 18 Spitzenplätze für HoreKa
- 20 KONKIT – ein Beitrag zur Kontrolle der Pandemie
- 22 WLAN für IoT-Geräte am KIT
- 24 Studenten-Meldeverfahren der gesetzlichen Krankenversicherungen
- 25 EPICUR Inter-University Campus

FORSCHUNG UND PROJEKTE

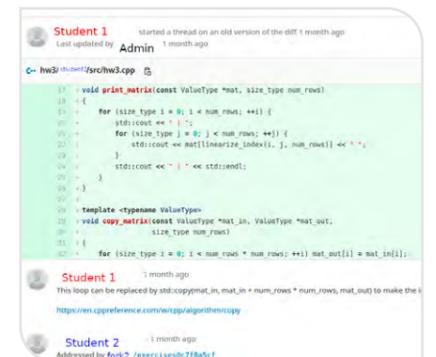
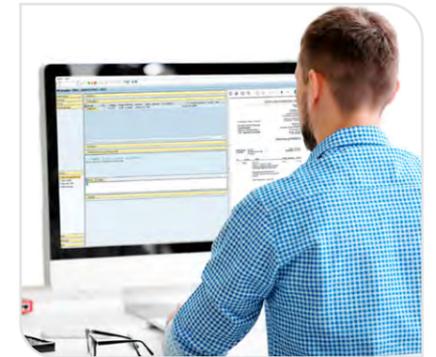
- 28 Transkription für frühneuzeitliche Drucke – vom Bilddigitalisat zum digitalen Volltext
- 30 Rückblick auf großartige Monate mit SORSE
- 32 Das EuroHPC-Projekt MICROCARD stellt sich vor
- 34 Maschinelles Lernen/Künstliche Intelligenz in Biosimulationen
- 34 Intelligentes Batteriemangement durch Spektroskopie und Maschinelles Lernen
- 35 wfGenes: Boosting productivity of workflow developers
- 38 Digitale Zwillinge aus dem virtuellen Materiallabor
- 39 Abschluss des Pilotprojekts Helmholtz Analytics Framework
- 42 O3as: ein Dienst zur Beurteilung des Ozonzustands

STUDIUM UND WISSENSVERMITTLUNG

- 44 Research Software Engineering in der Lehre

VERSCHIEDENES

- 46 Foto-Impressionen der HoreKa-Einweihung
- 47 Impressum & Kontakt



Schritt für Schritt hin zu SAP S/4HANA

In den kommenden Jahren wird schrittweise die bestehende SAP-Systemlandschaft aktualisiert. Dies beinhaltet den Wechsel zur Datenbanktechnologie SAP HANA, die Einführung der dazugehörigen Enterprise Resource Planning (ERP) Software der Firma SAP (vor allem S/4HANA) sowie die vermehrte Nutzung von Oberflächen im Fiori-Design. Die Aktualisierung besteht aus zahlreichen Einzelprojekten mit unterschiedlichen Zielen und Beteiligten.

Erste Projekte sind bereits in Umsetzung oder in Vorbereitung

Martin Hengel

SAP am KIT

Die SAP-Software stellt die zentrale Basis für die administrativen Prozesse im Finanzwesen, der Beschaffung, der Personaladministration u.v.m. am KIT. Schon in den Vorgängerinstitutionen des KIT war SAP im Einsatz – im Forschungszentrum bereits seit den 1980er Jahren. In den vergangenen zehn Jahren wurden die SAP-Systeme, z.B. mit Einführung der Employee Self Services (ESS) für die Beschäftigten des KIT, kontinuierlich weiterentwickelt. Auch in Zukunft wird die SAP-Software eine wichtige Säule für die digitale Unterstützung der Administration des KIT sein.

Neue Softwaregeneration basierend auf SAP HANA

In den letzten Jahren hat SAP sein Produktangebot aktualisiert und setzt

auf neue Technologien, insbesondere die Datenbank SAP HANA (kurz HANA). Bei HANA werden die Daten nicht mehr auf einer Festplatte gehalten, sondern direkt – in memory – im Arbeitsspeicher. Verbunden mit weiteren Techniken kann so eine deutlich gesteigerte Performance insbesondere bei der Verarbeitung großer Datenmengen erreicht werden.

Die neue Softwaregeneration von SAP ist auf HANA optimiert und wird zukünftig ausschließlich für diese Datenbank angeboten. Softwareprodukte wie S/4HANA oder BW/4HANA werden die bislang auch am KIT genutzten Systeme wie SAP ERP 6.0 oder BW (Business Warehouse) ablösen. Des Weiteren hat SAP in den letzten Jahren auch ein einheitliches Designkonzept unter dem Namen Fiori entwickelt (siehe SCC-News 1/2019). Fiori ermöglicht mittels rollenbasierter

Apps moderne, auch für Mobilgeräte geeignete, Benutzeroberflächen. Dies ist hauptsächlich für Anwendungen, die von einer hohen Anzahl an Personen genutzt werden, von großem Vorteil. Insbesondere die aktuell im SAP-Portal abgebildeten Funktionalitäten wie ESS sollen in Zukunft als Fiori-Apps zur Verfügung gestellt werden.

Vorgehen für die Aktualisierung

Die Wartung der derzeit am KIT genutzten SAP-Software läuft Ende 2027 aus. Schon in den letzten Jahren wurden erste Vorbereitungen (z.B. die Implementierung der „Neuen Hauptbuchhaltung“ im Finanzmodul) für die Einführung der neuen SAP-Softwaregeneration getroffen. Die Aktualisierung erfolgt jedoch nicht in Form eines großen Projekts, sondern wird schrittweise im Rahmen eines

mehrjährigen Programmes umgesetzt, das wiederum aus zahlreichen Einzelprojekten mit jeweils unterschiedlichen Zielen und Beteiligten besteht. Um dies zu strukturieren, wurden die geplante Zielsystemlandschaft für die Zeit ab 2027 (Abbildung 1) und die grundsätzliche Vorgehensweise in einer Roadmap festgehalten, die im Rahmen der IV-Governance des KIT im Lenkungsreis für IV-Themen der Administration vorgestellt und abgestimmt wurde.

Das KIT verfolgt dabei einen so genannten Konvertierungs- oder Brownfield-Ansatz, d.h. die bestehenden fachlichen Daten und Prozesse werden übernommen, und es wird anders als beim Greenfield-Ansatz kein komplett neues S/4HANA-System aufgebaut. Da SAP erst voraussichtlich Ende 2022 die S/4HANA-Module für die Personaladministration zur Verfügung stellen wird, ist eine vollständige Konver-

tionierung der bestehenden SAP-Systeme erst danach möglich. Die nächsten Jahre bis zur Konvertierung sollen genutzt werden, um grundlegende Vorarbeiten durchzuführen (z.B. die Einführung des Geschäftspartners als neues Stammdatum) und um einzelne Fachprozesse zu optimieren und anzupassen. Darunter fällt vor allem auch die Umsetzung zur Vollendung der Fusion des KIT in einem einheitlichen Rechtsrahmen¹. Des Weiteren sollen erste Projekte durchgeführt werden, die weitgehend unabhängig von der Konvertierung des SAP ERP sind. Hierzu zählt insbesondere die Ablösung von SRM (Supplier Relationship Management) und des SAP-Portals, für die SAP bereits die funktionale Weiterentwicklung eingestellt hat, durch neue, auf Fiori-Apps basierende Lösungen. Mit den Fiori-Apps wird sich die Benutzerfreundlichkeit der Oberflächen deutlich verbessern.

Erste Projekte sind bereits gestartet oder in Vorbereitung. So erfolgt aktuell ein Upgrade des Business Warehouse (BW) auf die Version 7.5 mit Umstellung der Datenbank auf HANA. Bereits mit dieser Umstellung sind wesentliche Verbesserungen bei der Performance zu erwarten. Des Weiteren kann das SCC als Betreiber Erfahrungen mit HANA-Datenbanken sammeln. Für die Ablösung des Portals existiert bereits ein Prototyp für ein Fiori-Launchpad (Abbildung 2), in dem in Zukunft die verschiedenen Funktionen wie ESS als Fiori-Apps zur Verfügung stehen werden. Und auch für die Ablösung des SRM werden derzeit in Workshops Fachkonzepte erarbeitet, in denen die Prozesse für und die Anforderungen an eine zukünftige Lösung festgehalten werden.

Damit sind aber lediglich die ersten Schritte einer langen Reise getan: Die umfassende Aktualisierung der SAP-Systemlandschaft wird für das SCC und die Administration des KIT in den kommenden Jahren eine zentrale Aufgabe sein.

Erste Projekte in Vorbereitung und Umsetzung

Die Aktualisierung der SAP-Systemlandschaft erfordert eine enge Zusammenarbeit zwischen den Dienstleistungseinheiten wie FIMA (Finanzmanagement), PSE (Personalservice), EVM (Einkauf, Verkauf und Materialwirtschaft) oder FM (Facility Management) und dem SCC. Viele Projekte werden einen technischen Fokus haben, in einzelnen Projekten müssen aber auch die Fachprozesse, die durch die SAP-Systeme unterstützt werden, betrachtet werden.

Step by step towards SAP S/4HANA

In the coming years, the existing SAP system environment will be updated step by step. This includes the switch to the SAP HANA database technology, the introduction of the associated Enterprise Resource Planning (ERP) software from SAP (primarily S/4HANA), and the increased use of user interfaces with Fiori design. The update consists of numerous individual projects with different goals and teams. The first projects are already being implemented or are in preparation.

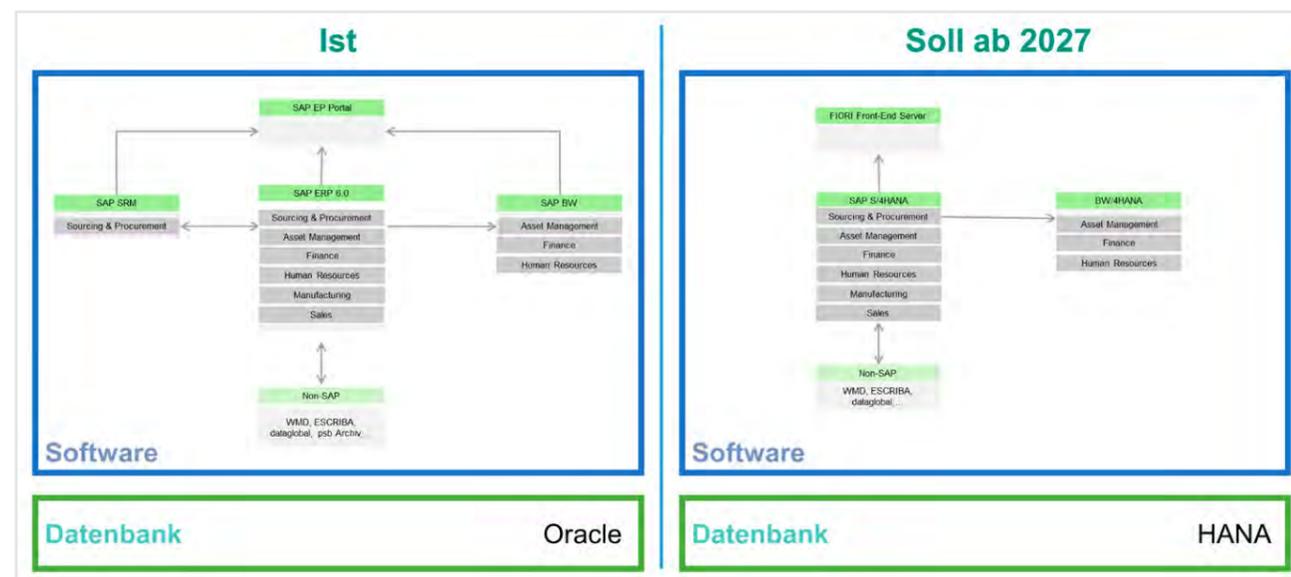


Abbildung 1: Übersicht aktuelle und geplante SAP-Systemlandschaft

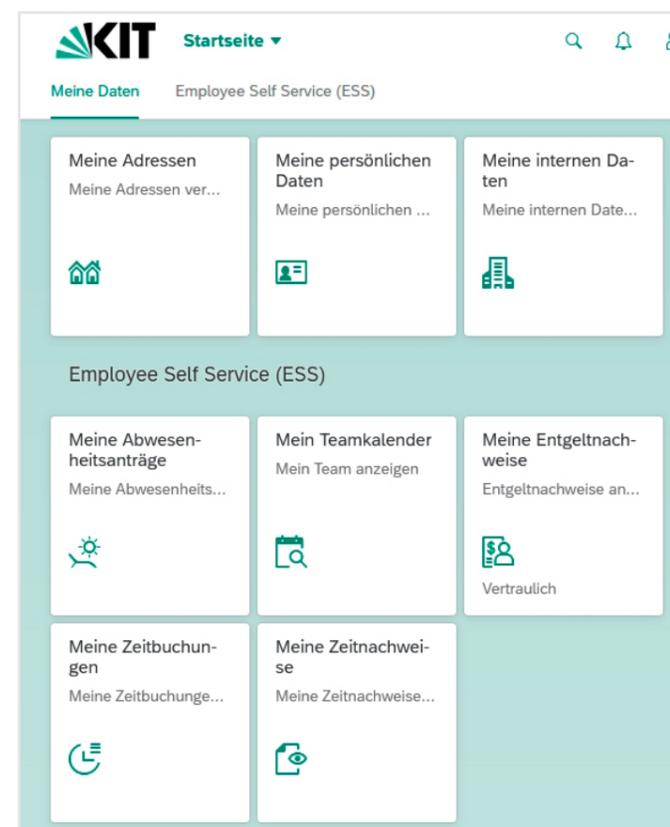


Abbildung 2: Prototyp Fiori-Launchpad

¹ www.kit.edu/kit/pi_2021_009_bund-und-land-einigen-sich-auf-weitere-schritte-zur-vollendung-der-fusion-am-kit.php

Einführung der elektronischen Ausgangsrechnung XRechnung

In Deutschland dürfen an Auftraggeber der öffentlichen Hand nur noch elektronische Rechnungen übermittelt werden. Das Finanzmanagement des KIT und das SCC haben daher das SAP-ERP-System mit einer Softwarelösung zur Erstellung von elektronischen Ausgangsrechnungen im Format XRechnung erweitert.

Gerald Helck

Im Zeitfenster Oktober 2020 bis März 2021 haben FIMA und SCC eine Lösung für den Versand elektronischer Ausgangsrechnungen im Format XRechnung implementiert. Alle rechnungsrelevanten Daten werden hier in Form einer strukturierten XML-Datei gehalten. Anders als bei einer PDF-Datei haben diese Daten keine bildgebenden Teile.

Die XRechnung ist der umgesetzte Standard der deutschen e-Rechnungsverordnung¹, die wiederum auf der EU-Richtlinie 2014/55/EU beruht. Die Einführung der Software wurde notwendig, weil in Deutschland nur noch elektronische Rechnungen an Auftraggeber der öffentlichen Hand übermittelt werden dürfen. Rechnungen in Papierform oder als PDF-Datei werden in der Regel von diesen Kunden nicht mehr akzeptiert.

Die XRechnung ermöglicht einen vollautomatischen Datenaustausch zwischen Rechnungssender und -empfänger. Mit der Nutzung einer elektronischen Ausgangsrechnung lassen sich deutliche Kostensenkungen erreichen.

Da auf der Empfängerseite der gesamte Prozess des Einlesens und Scannens, des Validierens und Korrigierens entfällt, werden Fehler minimiert und Bearbeitungszeiten wesentlich reduziert. Für die Generierung und den Versand von XRechnungen wurde AFI Billing Control als Lösung ausgewählt. Diese ist vollständig in das SAP-System integriert und steht nicht nur für die aktuell genutzte SAP-ERP-Lösung zur Verfügung, sondern ist auch im Rahmen einer künftigen SAP S/4HANA-Lösung (siehe auch S. 4) nutzbar.

Der Versand der Rechnungen zum Kunden erfolgt über das PEPPOL-Netzwerk (Pan-European Public Procurement OnLine)². Rechnungssteller sowie Rechnungsempfänger erhalten hierbei über einen sogenannten Access Point Zugang zur PEPPOL-Infrastruktur.

Neben der Angabe des Access Points ist auch die sogenannte Leitweg-ID zu hinterlegen. Die Leitweg-ID ist ein eindeutiger Schlüssel für einen öffentlichen Auftraggeber. Nur die kombinierte Angabe beider Informationen gewährleistet eine erfolgreiche Weiterleitung der Rechnung. Im Rahmen des Einführungsprojekts wurde für alle Kunden, die an der Nutzung der elektronischen Ausgangsrechnung interessiert waren, die Pflege dieser Merkmale in den entsprechenden Stammsätzen durchgeführt.

Die Produktivsetzung der elektronischen Ausgangsrechnung erfolgte im März 2021 und verlief ohne nennenswerte Probleme.

Adoption of the Electronic Outgoing Invoice XRechnung

In Germany, only electronic invoices may be sent to clients in the public sector. KIT's financial management and SCC have therefore extended the SAP ERP system with a software solution for creating electronic outgoing invoices in XRechnung format.

¹ www.bmi.bund.de/SharedDocs/downloads/DE/gesetzestexte/e-rechnungsverordnung.pdf

² peppol.eu/what-is-peppol

Aktionstag zum European Cyber Security Month 2020 am KIT

Das KIT beteiligte sich auch dieses Jahr im Rahmen des European Cyber Security Month mit vielfältigen Angeboten und Materialien, um Beschäftigte und Studierende des KIT für die Informationssicherheit zu sensibilisieren. Der Fokus lag auf der Informationssicherheit allgemein und der Erkennung betrügerischer Nachrichten sowie dem sicheren Umgang mit diesen.

Juris Hartwig

Der European Cyber Security Month (ECSM) sensibilisiert seit 2012 europaweit Bürgerinnen und Bürger sowie Organisationen für den umsichtigen und verantwortungsbewussten Umgang im Cyber-Raum. Als Pilotprojekt gestartet, ist der Aktionsmonat seit dem Jahr 2013 ein regelmäßig stattfindendes europaweites Format.

Unter Federführung der europäischen IT-Sicherheitsbehörde ENISA bieten die Mitgliedsstaaten der Europäischen Union während des ECSM verschiedene Veranstaltungen und Informationen zum Thema Cyber-Sicherheit an.

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) informiert als Koordinierungsstelle für den ECSM in Deutschland Organisationen und Unternehmen, die sich mit Fragen der IT-Sicherheit befassen, und wirbt dafür, sich entsprechend mit eigenen Aktionen einzubringen.

Beiträge zum ECSM am KIT

Auch das KIT beteiligte sich dieses Jahr im Rahmen des ECSM mit vielfältigen Angeboten und Materialien zu wichtigen

Aspekten aus der Informationssicherheit. Der Fokus lag auf der IT-Sicherheit (neu: Informationssicherheit) allgemein und der Erkennung betrügerischer Nachrichten sowie dem sicheren Umgang mit diesen. Das Hauptereignis des ECSM fand auf Grund der Corona-Situation am 18. November 2020 zum ersten Mal im Onlineformat via Microsoft Teams statt. Die Beschäftigten und Studierenden hatten die Möglichkeit, sich umfassend zu informieren und Fragen zu stellen.

Neue, wissenschaftlich geprüfte Materialien

Die Awareness-Materialien werden von der Forschungsgruppe SECUSO (Security • Usability • Society)¹, vom Steinbuch Centre for Computing (SCC) und vom Informationssicherheitsbeauftragten (ISB)² bereitgestellt. Neu sind dieses Jahr die Erklärvideos zur Informationssicherheit, die auf der Webseite des Informationssicherheitsbeauftragten³ veröffentlicht wurden: Wie erkenne ich betrügerische Nachrichten oder wie schütze ich meine Informationen unterwegs, wenn ich auf einer Geschäftsreise bin, worauf soll ich bei der Nutzung von Cloud-Diensten oder

Sozialen Netzwerken achten? Auf diese und andere Fragen findet jede und jeder eine Antwort in den dort hinterlegten Erklärvideos und Materialien.

Fazit

Aufgrund der recht positiven und hohen Resonanz bei den Beschäftigten des KIT werden auch künftig solche Aktionstage durchgeführt. Dazu gehören neben den Aktionen im ECSM auch der Internationale Safer Internet Day⁴.

Day of Action for European Cyber Security Month

European Cyber Security Month 2020 took place on November 18th 2020 in the new online format (Microsoft Teams). The focus this year was on IT security (new: information security) in general and the detection of fraudulent messages and the secure handling of them.



¹ secuso.aifb.kit.edu

² www.isb.kit.edu

³ www.klicksafe.de/ueber-klicksafe/safer-internet-day

⁴ www.isb.kit.edu/awareness-materialien.php

Bis zum GridKa und noch viel weiter – Opportunistische Ressourcen mit COBaLD/TARDIS

Als Mitglied im Weltweiten LHC Computing Grid ist das SCC und insbesondere das Grid Computing Centre Karlsruhe, kurz GridKa, ein langjähriger Partner der internationalen Forschungsgemeinschaft der Hochenergie- und Astroteilchenphysik. Neben den offensichtlichen Netzwerk-, Speicher- und Rechenressourcen in den Schränken hat das GridKa noch sehr viel mehr zu bieten. Dazu gehört jede Menge Know-How, Community-Arbeit, Entwicklungsleistung und Erfahrung im Dienstbetrieb. Eine der aktuellen Entwicklungen ist die COBaLD/TARDIS Software Suite, mit der Rechenressourcen aus Grid-, Cloud-, HPC- und anderen Quellen als einheitlicher, virtueller Rechencluster transparent zusammengeführt und für die Wissenschaft bereitgestellt werden können.

Max Fischer, Eileen Kühn, Manuel Giffels

Die Nutzung modernster Rechenressourcen ist ein Standbein der heutigen Wissenschaft. Dementsprechend breit ist auch die Auswahl: rasend schnelle High-Performance Computer, unermüdliche High-Throughput Rechencluster, maßgeschneiderte Cloud-Ressourcen oder mit Herzblut umsorgte Institutsrechner – was immer das Herz begehrt. Die Verfügbarkeit einer breiten Auswahl spezialisierter Ressourcen allein erlaubt aber noch nicht deren effiziente Nutzung. Stattdessen verlangt dies nicht nur viel technische Expertise für die Entscheidungsfindung, sondern teilweise auch umfangreiche Entwicklungen zur Steuerung und Anpassung von Rechenaufgaben. Dies bindet im Umkehrschluss Forschungsgruppen meist auch an spezifische Systeme. Dabei möchte und muss man doch einfach nur "Wissenschaft machen", ganz unabhängig von der konkreten Wahl der Ressourcen.

weit arbeiten können, standardisiert das Weltweite LHC Computing Grid (WLCG)¹ den Zugriff auf Speicher- und Rechenressourcen. Neben einheitlichen Services, Protokollen und Authentifizierungsverfahren kommen für Rechenressourcen sogenannte Overlay Batch Systeme (OBS) zum Einsatz: Diese virtuellen Rechencluster buchen automatisch Rechenzeit bei Rechenzentren und stellen ihren Nutzern die ihnen zugewiesenen Ressourcen als – scheinbar – einheitliches System bereit.

Von den Großen lernen

Zugleich sind diese OBS-Systeme extrem skalierbar: So hat die CMS-Kollaboration am CERN bereits vor Jahren gezeigt, dass ein OBS mit einer Million CPU-Kernen problemlos realisierbar ist.² Gleichzeitig ist auch die Benutzbarkeit "skalierbar": Endnutzerinnen und Endnutzer arbeiten nur mit dem OBS und erhalten damit Zugang zu hunderten zugrundeliegender Batch-Systeme, während die Administrierenden der Rechencluster lediglich mit einer Handvoll Kollaborationen zusammenarbeiten und damit automatisch tausende Endnutzerinnen und Endnutzer versorgen können.

Geht ab wie eine Rakete!

Die Forschung und Entwicklung zu opportunistischen Ressourcen hat nicht erst 2018 mit COBaLD/TARDIS begonnen, sondern baut auf einer langen Entwicklungshistorie am KIT auf. Bereits 2007 begannen die ersten Entwicklungsarbeiten am Institut für Experimentelle Teilchenphysik (ETP) zur dynamischen Nutzung von Virtualisierung, um Rechenressourcen nutzbar machen zu können, welche nicht primär für die Hochenergiephysik vorgesehen waren. Seit 2010 stand der am ETP entwickelte Meta-Scheduler ROCED (Responsive On-Demand Cloud-enabled Deployment) im Zentrum der Forschung zur Verwendung von Virtualisierung und Containern für die opportunistische Nutzung von Rechenressourcen. Während der mehrjährigen Entwicklungs- und Nutzungsdauer von ROCED wurden sowohl kommerzielle Anbieter wie Amazon/AWS und 1&1 als auch wissenschaftliche Anbieter wie der NEMO-Cluster in Freiburg und der ForHLR II-Cluster am SCC eingebunden. Als wissenschaftliche Forschungssoftware brachte ROCED viele Erkenntnisse, wie opportunistische Ressourcen effizient genutzt werden sollten. Auf dieser Basis wurde 2018 mit der prototypischen Entwicklung der Applikationen COBaLD und TARDIS begonnen. Seit 2019 ist die COBaLD/TARDIS Software Suite der offizielle Nachfolger von ROCED.

Damit ein OBS jederzeit den Bedarf seiner Nutzenden abdecken kann, kommen sogenannte Meta-Scheduler zum Einsatz. Statt wie normale Scheduler eines Batch-Systems Rechenjobs an verfügbare Ressourcen zuzuweisen, entscheidet ein Meta-Scheduler, von welchen Anbietern Ressourcen gebucht werden sollten, da-

Bonn über Bonn in Bonn

Die Etablierung von COBaLD/TARDIS wäre nicht möglich gewesen ohne die hervorragende Kooperation mit der Universität Bonn. Die Universität Bonn ist praktisch der erste externe Partner, der COBaLD/TARDIS lokal und eigenständig betreibt, um vollkommen autonom Rechenressourcen in das Overlay Batch System am GridKa zu integrieren. In Bonn wird COBaLD/TARDIS von den lokalen Expertinnen und Experten eingesetzt, um einerseits Ressourcen des Rechenclusters des Physikalischen Instituts opportunistisch für die ATLAS- und Belle2-Kollaborationen bereitzustellen und andererseits auch den HPC-Cluster bonna anzubinden und zu betreuen. Durch die Integration in das GridKa OBS sind alle Ressourcen in Bonn automatisch durch das WLCG nutzbar, ohne dass die Universität Bonn eigene Grid Services betreiben muss. Die Zusammenarbeit mit der Universität Bonn war nicht nur ein unersetzlicher Feldversuch, der so manchen Bug aufdeckte, sondern sie war und ist zudem eine wichtige Stütze bei der Weiterentwicklung von COBaLD/TARDIS hinsichtlich Nutzbarkeit, Funktionalität und Robustheit.

mit der eigentliche Scheduler überhaupt genug geeignete Ressourcen zur Verfügung hat. Durch die Standardisierung von Rechenjobs und Services im WLCG ist es dabei für einen Meta-Scheduler ein Leichtes, den Ressourcenbedarf präzise genug vorherzusagen.

In freier Wildbahn

Wer sich die Rechenressourcen des SCC genauer anschaut, weiß jedoch: Das WLCG ist längst nicht alles. Für viele Forschende unterschiedlichster Domänen ist beispielsweise der Parallelrechner bzw UniCluster unentbehrlich, und auch der Supercomputer HoreKa ist auf sehr viel mehr Anwendungsfälle, als die des WLCG, ausgelegt. Jedes dieser Systeme hat unterschiedliche Zugangsmethodiken, Ressourcenprofile oder Wartezeiten. Auch das GridKa selbst betreibt für lokale Nutzergruppen mit dem TOpAS-Cluster

spezielle Hardware, wie Grafikbeschleuniger (GPUs), und entspricht somit auch nicht der Standardisierung im WLCG. Die am SCC zur Verfügung stehenden Ressourcen reichen bereits aus, um die existierenden Vorhersagemethoden zu Ressourcenverfügbarkeit und -bedarf und damit auch die Meta-Scheduler der Hochenergiephysik an ihre Grenzen zu bringen.

Um nun das Konzept eines OBS auch über viele heterogene Ressourcen zu ermöglichen, verfolgt eine F&E-Gruppe, eine Kooperation des Instituts für Experimentelle Teilchenphysik (ETP) und des SCC, mit der COBaLD/TARDIS Software Suite einen anderen Ansatz. Die zugrundeliegende Idee dieser Entwicklungen ist, dass Vorhersagen in solch komplexen Systemen weder skalierbar noch praktikabel sind. Statt vorherzusagen, wie die Bedarfe der Nutzenden, Scheduler und

Ressourcen zusammenspielen, basiert die Umsetzung darauf, verfügbare Ressourcen zu überwachen – gut ausgelastete Ressourcen sind offensichtlich aus Sicht der Nutzenden und Scheduler nützlich und werden daher von COBaLD/TARDIS hinzu gebucht, schlecht genutzte Ressourcen werden stattdessen von COBaLD/TARDIS freigegeben. Durch diesen reaktiven Ansatz können praktisch beliebig viele Ressourcen ohne präzises Wissen über Nutzeranforderungen, Scheduling-Strategien oder Ressourcencharakteristika verwaltet werden.

COBaLD/TARDIS dient nicht nur als Plattform zur Forschung über Meta-Scheduler, sondern ist gleichzeitig auch ein einsatzbereites und vielfach erprobtes Produkt. Bereits kurz nach Entwicklungsbeginn im Jahr 2018 wurde COBaLD/TARDIS zur dynamischen Bereitstellung von Rechenressourcen aus der Helix Nebula Science

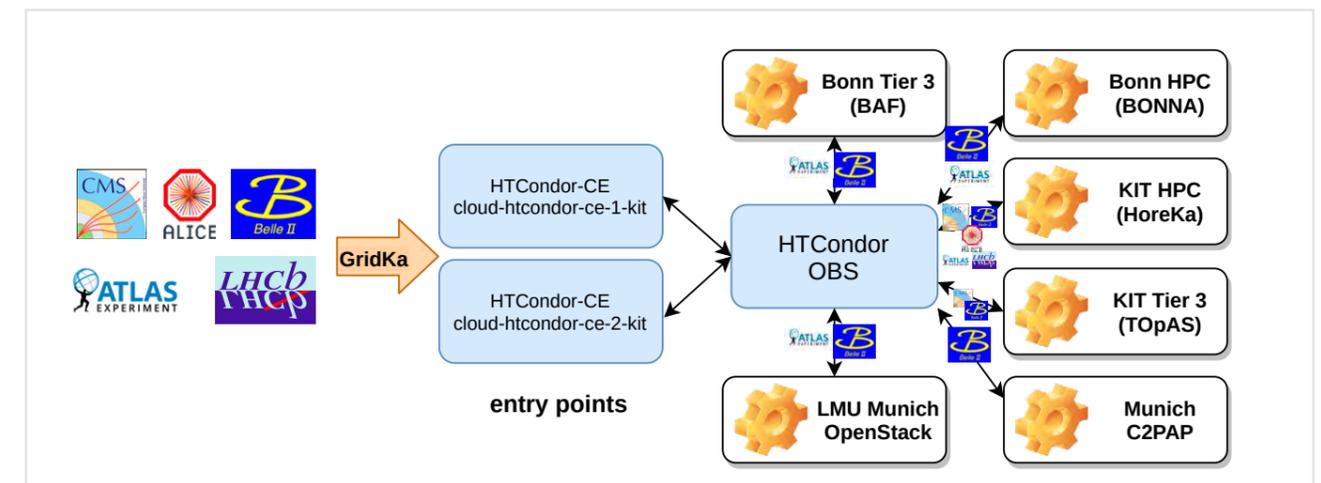


Abbildung 1: Setup des GridKa Overlay Batch Systems (OBS) und kooperierender Ressourcenbetreiber.

Cloud – ein Wegbereiter der European Open Science Cloud – für GridKa und seiner Wissenschaftsgemeinschaften eingesetzt. Seit diesem ersten Feldtest wurde nicht nur ein immenser Erfahrungsschatz zu den vielfältigen Einsatzmöglichkeiten dieser Form des Meta-Schedulings aufgebaut, sondern auch eine überaus robuste Software, die es ermöglicht, eine Vielzahl an Ressourcenprovidern zu einem einzigen Ressourcenpool zusammenzubinden. Auch wenn COBalD/TARDIS spezifisch für Anwendungsfälle der HEP entwickelt wurde, ist die zentrale Idee – Reaktion ohne Systemwissen – nicht auf diesen Anwendungsfall begrenzt und ist stattdessen für beliebige ähnliche Anwendungsfälle denk- und nutzbar.

Viele Stücke vom Kuchen

Wie sich herausgestellt hat, macht ein reaktiver Ansatz das Meta-Scheduling unterschiedlicher Ressourcen nicht nur möglich, sondern auch praktisch. Ohne den Anspruch alle Interaktionen vorherzusagen, können einzelne Ressourcenbetreiber isoliert voneinander gehandhabt werden. Dadurch kann mit COBalD/TARDIS ein einziges OBS aufgebaut werden, das gleichzeitig unterschiedliche, nachfolgend beschriebene Nutzungsmo-

delle unterstützt: Ressourcenbetreiber können ihre Ressourcen selbst bereitstellen, beispielsweise um mit Backfilling vorübergehend ungenutzte Ressourcen auszulasten. Über Anträge eingeworbene Rechenzeit auf Rechenclustern kann nach Bedarf abgerufen werden, und gleichzeitig können die Rechenaufgaben optimal über mehrere Cluster verteilt werden. Isolierte Rechenressourcen, wie beispielsweise ein "Institutscluster im Keller", können größeren Infrastrukturen beitreten. Auch kurzfristig verfügbare Ressourcen, wie Cloud-Ressourcen, können transparent zu anderen hinzugeschaltet werden.

Daher verwundert es nicht, dass COBalD/TARDIS heute genau dafür genutzt wird: Das GridKa betreibt zusammen mit mehreren internen und externen Partnern einen virtuellen Cluster, über den der HEP-Gemeinschaft unterschiedlichste Rechenressourcen zur Verfügung gestellt werden. Dabei kommen mehrere separate Instanzen von COBalD/TARDIS zum Einsatz, etwa zum Backfilling auf den HPC-Clustern des SCC und zur Einbindung spezieller Hardware wie GPUs auf dem TOpAS, aber auch zur Nutzung externer Instituts- und HPC-Cluster (Abbildung 1). Zusätzlich tut GridKa das, was es ohnehin gut kann (SCC-News 2/2020,

S. 12): ein Batch System betreiben – als OBS statt für lokale Ressourcen – das über sogenannte Compute Elemente ins WLCG eingegliedert wird.

Dabei steht an praktisch jedem Punkt die Nutzbarkeit im Mittelpunkt. Für Ressourcenbetreiber ergibt sich eine einfache Möglichkeit, ihre Ressourcen einer breiten Nutzerschaft zugänglich zu machen und hohe Auslastungen bei effizienter Nutzung zu erreichen. Für die Betreiber des OBS ist es ein Leichtes, sowohl die Ressourcenbetreiber als auch die Mitglieder ihrer Wissenschaftsgemeinschaft zu unterstützen. Endnutzerinnen und Endnutzer erhalten unkompliziert Zugriff auf eine große Auswahl und Anzahl unterschiedlicher Ressourcen.

Was die Zukunft bringt...

Mit der COBalD/TARDIS Software Suite hat die F&E-Gruppe des ETP und SCC bereits einiges ermöglicht und erforscht. Nach dem Vorbild des GridKa OBS, das mehrere kooperierende Wissenschaftszentren und Ressourcenbetreiber vereint, können auch in anderen Regionen und für andere Wissenschaftszweige vergleichbare OBS geschaffen und eingesetzt werden. Die COBalD/TARDIS Software Suite ist zwar von sich aus schon weitaus weniger komplex als vergleichbare Meta-Scheduler, trotzdem steht weiterhin eine bessere Nutzbarkeit im Mittelpunkt: Die Erfahrungen aus dem GridKa OBS werden stetig in Form von Vorlagen, beispielsweise für Konfigurationen oder Container, zusammengefasst und bereitgestellt und können mit wenig Aufwand selbst genutzt werden. Das Team widmet sich aber auch aktuellen Herausforderungen, die erst durch die neue Dimension der Ressourcen- und Betreiberföderation entstehen – beispielsweise dem Accounting und der Attribution der vielen Ressourcenbetreiber oder der Vereinbarkeit anonymer Nutzergruppen eines OBS mit auf Einzelnutzern ausgelegten Nutzungsrichtlinien.

³ doi.org/10.5281/zenodo.848818

Und auch weiterhin dreht sich vieles um das, womit alles ursprünglich angefangen hat: In Zusammenarbeit mit dem HPC-Team des SCC wurden vor kurzem die Rechenressourcen des neuen HoreKa-Clusters während dessen Inbetriebnahme durch COBalD/TARDIS

für einen Skalierungstest benutzt. Über einen Zeitraum von einer Woche konnten damit über das OBS des GridKa mehr als 12.000 zusätzliche Rechenkern für das CMS-Experiment am CERN zur Verfügung gestellt werden (Abbildung 2). Das sind ungefähr dreimal so

viele Rechenressourcen, wie dem CMS-Experiment am GridKa statisch zur Verfügung stehen. Insgesamt wurde damit in dieser einen Woche eine Rechenzeit von knapp 2,2 Millionen Core-Stunden durch HoreKa und COBalD/TARDIS nutzbar gemacht.

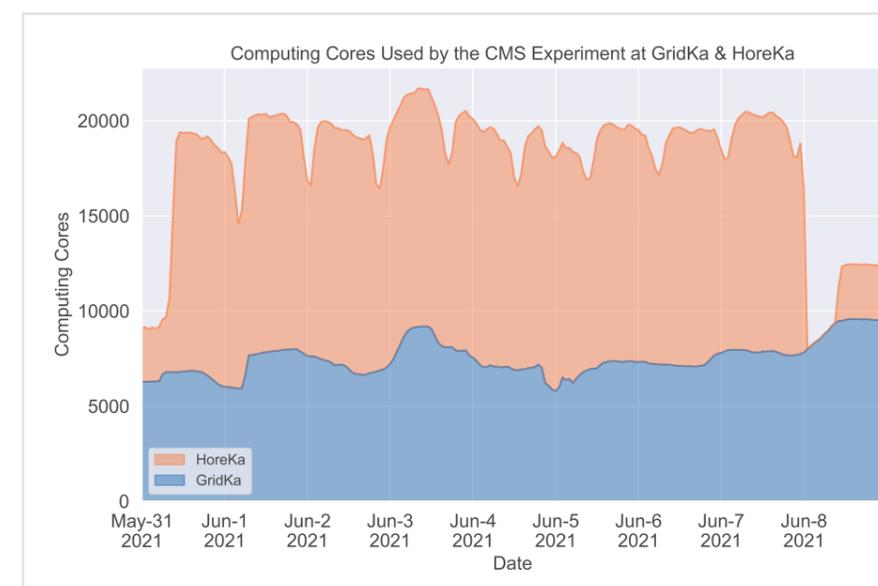


Abbildung 2: Ressourcen für das CMS-Experiment während des Skalierungstests des HoreKa-Clusters

Kontakt



<https://twitter.com/matterminers>



<https://gitter.im/MatterMiners/community>



<https://matterminers.github.io>



matterminers@lists.kit.edu

Beyond GridKa: Opportunistic Resource usage via COBalD/TARDIS

As a member of the Worldwide LHC Computing Grid, SCC and especially GridKa is a longstanding partner of the High Energy Physics and Astro Particle Physics communities. In addition to the obvious contribution of compute and storage resources, GridKa provides significant know-how, community involvement, and development efforts. Recently, the COBalD/TARDIS Software Suite has been developed to accumulate compute resources from HPC, Cloud, Grid, and other providers into a single, virtual compute cluster.

The opportunistic usage of such resources, that is the temporary acquisition only while there is demand, has proven to be extremely challenging even for moderately heterogeneous resources. The novel approach of COBalD/TARDIS is to tackle this problem not with a more complex but a much simpler approach: Instead of predicting which resources will be needed, the system is merely reacting to observed needs. In practice, this foregoes an unachievable optimal solution in favour of an effective, robust, and scalable heuristic.

The simple approach of COBalD/TARDIS has proven to enable the federation of independent resource providers into a single Overlay Batch System. GridKa itself provides such an Overlay Batch System, integrating the resources of several cooperating HPC centers and university clusters across Germany. The most recent success is the opportunistic usage of the SCC HoreKa HPC Cluster during its burn-in phase – using on average about 12,000 CPU cores over one week for a total of roughly 2.2 million core hours.

Innovative Multifaktor-Authentifizierung für sicheren HPC-Zugang

Im Frühjahr 2020 wurden Cyberattacken auf mehrere HPC-Systeme in Deutschland und im internationalen Umfeld bekannt. Unbekannte hatten sich mit Hilfe von gestohlenen SSH-Schlüsseln Zugang zu den Systemen verschafft. Mit der Einführung einer Multifaktor-Authentifizierung und einer dezentralen Speicherung der SSH-Schlüssel konnte auf innovative und zukunftssichere Weise das Sicherheitsniveau der HPC-Systeme des SCC signifikant erhöht werden, während die notwendigen Änderungen auf ein Minimum beschränkt blieben.

Jennifer Buchmüller, Simon Raffener, Michael Simon, Ulrich Weiß

Im internationalen Wissenschaftsbetrieb ist es durchaus üblich, dass Forschende gleichzeitig Zugang zu mehreren unterschiedlichen HPC-Systemen haben. Die Nutzung von sogenannten SSH-Keys, speziellen kryptographischen Schlüsselpaaren, vereinfacht dabei den Zugang und auch die Automatisierung häufiger Abläufe.

Viele HPC-User benutzten den gleichen SSH-Schlüssel für den Zugang zu mehreren HPC-Systemen. Dies ist zwar sehr komfortabel, stellt aber auch ein willkommenes Ziel für Cyberattacken dar. Einer unbekannt Gruppe von Angreifenden gelang es im Zeitraum September 2019 bis ca. März 2020, die eigentlich geheimen privaten Schlüssel von den Systemen einiger User abzugreifen und sich dadurch Zugang zu einer größeren Zahl von Hochleistungsrechnern zu verschaffen.

Nachdem sich die Angreifer unberechtigt auf einem System angemeldet hatten, beschafften sie sich Administratorrechte, suchten gezielt nach weiteren SSH-Keys sowie beispielsweise auch Angaben aus Log-Dateien, welche auf weitere verwundbare HPC-Systeme hinweisen könnten. Gleichzeitig modifizierten sie Betriebssystemdienste, löschten Systemprotokolle und installierten eine sogenannte Backdoor (Hintertür), um auch weiterhin Zugriff auf das System zu haben.

Wie genau sich die Angreifer Administratorrechte beschaffen konnten, ist bis heute nicht abschließend geklärt. Erst

nach mehreren Wochen konnten auf zwei nicht vom SCC betriebenen HPC-Systemen Überreste von Schadsoftware – sogenannten Exploits – entdeckt werden, mit denen unter Ausnutzung von bislang unbekannt Sicherheitslücken eine Privilegienweiterung möglich war.

Erhöhung des Sicherheitsniveaus

Als Reaktion auf diese Angriffe wurden am SCC in Zusammenarbeit mit internationalen Partnern Maßnahmen ergriffen, um die Systeme zukünftig noch besser vor Angriffen dieser Art zu schützen. Zu diesen Maßnahmen gehören insbesondere die Anpassung der Sicherheitsrichtlinien, Multifaktor-Authentifizierungsverfahren, die dezentrale Speicherung von SSH-Keys außerhalb der HPC-Systeme, eine Ausweitung des Monitorings, die Einschränkung des Netzverkehrs, aber in zweiter Linie auch eine stärkere Sensibilisierung der Nutzenden für den sicheren Umgang mit ihren Zugangsdaten.

Bei all diesen Maßnahmen war jeweils eine Abwägung erforderlich, um das höchstmögliche Sicherheitsniveau zu erreichen und gleichzeitig die erforderliche flexible Arbeit in der Wissenschaft nicht zu behindern. Insbesondere bei den SSH-Keys erwies sich dies als nicht trivial. Deren Schwachstellen waren zwar bekannt, eine pauschale Deaktivierung oder die unbedachte Einführung einer zusätzlichen Multi-Faktor-Authentifizierung hätten jedoch die Hürden für die Automatisierung von Abläufen dermaßen erhöht, dass eine

sinnvolle Nutzung nicht mehr möglich gewesen wäre.

RegApp – Authentifizierungs- und Autorisierungs-Infrastruktur

Im Gegensatz zu vielen anderen HPC-Betriebsstandorten waren die Systeme am SCC bereits an ein leistungsfähiges, föderiertes Identitätsmanagement (bWIDM) auf Basis der sogenannten RegApp angebunden. Dabei handelt es sich um ein föderiertes Open-Source-Identity-Management-System, das maßgeblich am SCC entwickelt wird¹ und landesweit an Universitäten und Hochschulen zum Einsatz kommt.

Die Implementierung der für die Absicherung der HPC-Systeme notwendigen Funktionen in der RegApp konnte innerhalb weniger Wochen abgeschlossen werden. Die HPC-Systembetreiber übernahmen deren Anbindung an die neuen Schnittstellen, entwickelten ein neuartiges Verfahren für die sichere Automatisierung wissenschaftlicher Abläufe und wickelten die umfangreiche Kommunikation mit den Nutzenden ab.

Neue Oberfläche und Schnittstellen

Im Detail wurde innerhalb der RegApp die Benutzeroberfläche dahingehend erweitert, dass Nutzende nun den öffentlichen Teil ihrer SSH-Keys und Smartphone- bzw. Hardware-Token über einen Webbrowser verwalten können (Abbil-

dung 1). Gleichzeitig wurden im Backend neue Schnittstellen für die Abfrage dieser Daten durch die HPC-Systeme geschaffen. Zudem wurde eine Reihe von Regeln implementiert, welche die Handhabung von Dienstpasswörtern, Einmalpasswörtern und SSH-Keys so miteinander kombiniert, dass Passwörter nicht pauschal bei jedem Zugriff eingegeben werden müssen.

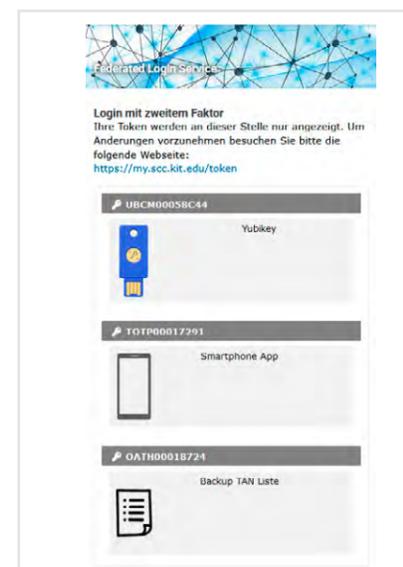


Abbildung 1: Token-Verwaltung in der RegApp

Die Implementierung dieser Funktionen in der RegApp bietet gleich mehrere Vorteile. Zum einen werden SSH-Keys nun nicht mehr auf dem HPC-System hinterlegt, was einen deutlichen Sicherheitsgewinn mit sich bringt. Zum anderen müssen auf Seiten des HPC-Systems nur die dort sowieso schon vorhandenen Standardkomponenten (OpenSSH-Server) neu konfiguriert werden. Der Einsatz von Patches oder gar proprietären Erweiterungen ist nicht erforderlich (Abbildung 2).

Neben dem KIT haben bereits mehrere andere Standorte eine auf die RegApp aufbauende Zwei-Faktor-Authentifizierungslösung für die eigenen HPC-Systeme umgesetzt. Zudem bietet das SCC anderen bWHP-Standorten² den zentralen Betrieb der RegApp als Dienstleistung an (Abbildung 3). Der Einsatz der Software

ist auch im Verbund des Nationalen Hochleistungsrechners (NHR) geplant und wird im Rahmen eines Projekts umgesetzt.

Bewährte Technik im Flächeneinsatz

Derzeit verwaltet die am SCC betriebene RegApp im bWIDM-Umfeld mehr als 38.000 Nutzerinnen und Nutzer, von denen 3.200 für die angeschlossenen HPC-Systeme registriert sind. Letztere führen täglich im Durchschnitt über 1.200 Logins mit dem zweiten Faktor auf den HPC-Systemen aus.

Zusätzlich zur direkten Provisionierung von Verzeichnisdiensten (z.B. LDAP) kann die RegApp auch als IdP/SP-Proxy fungieren. Dabei kombiniert der Proxy einen nach außen gerichteten Service Provider (SP), der sich in eine multilaterale Föderation wie z.B. die DFN-AAI oder die Helmholtz AAI integriert, mit einem nach innen gerichteten Identity Provider (IdP), der den Diensten die benötigten Benutzerattribute bereitstellt. Die RegApp bietet als Proxy komplexe Funktionalitäten wie Identitätsverknüpfung (z.B. User Accounts bei unterschiedlichen Einrichtungen), Attributanreicherung (z.B. für Gruppen- und Community-Verwaltungen) und Protokollübersetzung (z.B. Anbindung eines OpenID-Connect-basierten Dienstes an SAML-Infrastrukturen).

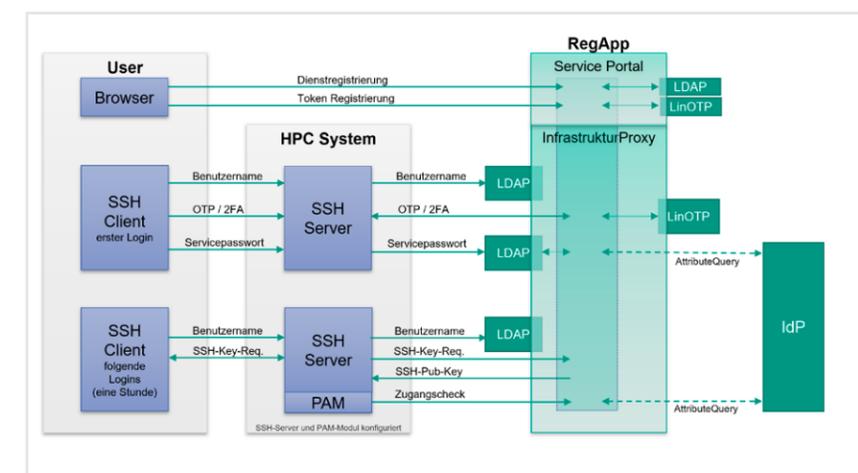


Abbildung 2: Schematischer Ablauf von Benutzerregistrierung mittels RegApp, initialem SSH-Login mit notwendiger Zwei-Faktor-Authentifizierung und nachfolgend automatischen Logins durch SSH-Keys

Am KIT erfolgt der Einsatz als IdP/SP-Proxy beispielsweise, um den Zugang auf die HPC-Systeme mit JupyterHub und Gitlab zu kombinieren. Auch bei auf OpenID Connect basierenden Diensten kann ein Login mit hinterlegtem zweiten Faktor erfolgen.

Weiterentwicklung der RegApp

Die Zukunftspläne sehen unter anderem die Realisierung einer Verwaltung von Projektgruppen vor. Eine Projektgruppe ist ein virtueller Verbund von Benutzeridentitäten bzw. -konten, die nicht unbedingt zur gleichen Heimatorganisation gehören müssen. Die Mitgliedschaft in einer solchen Gruppe stellt die Grundlage für Autorisierungsentscheidungen dar, z.B. den Zugang zu einem Dienst oder den Zugriff auf Daten. Solche Mitgliedschaften müssen an den entsprechenden Dienst oder die Community weitergegeben werden. Allerdings ist – insbesondere in Hinblick auf Communities – nicht jeder Dienst in der Lage, selbst eine entsprechende Benutzerautorisierung durchzuführen. Die RegApp kann diese daher übersetzen und in Form von bereits unterstützten Attributen ausliefern.

Ein besonderer Fokus liegt auch auf der Implementierung der mit Projektgruppen und Communities verbundenen Prozesse,

¹ www.scc.kit.edu/dienste/regapp

² www.bwhpc.de

wie z.B. der Provisionierung und Deprovisionierung von Projektgruppen sowie dem Entwurf und der Implementierung eines sinnvollen Rechte-/Rollenkonzepts zur Verwaltung der Gruppen bzw. ihrer Mitglieder.

Diesem Ansatz folgend wird die Implementierung eine erweiterte Dienstintegration durch moderne Authentifizierungsmethoden und Protokolle ermöglichen. Dies gewährleistet die Offenheit des

Systems und die zukunftssichere Integration weiterer Datendienste. Die Etablierung geeigneter Account-Verknüpfungsprozeduren innerhalb der RegApp ermöglicht die Zuordnung eindeutiger Identitäten auf konsistente und einheitliche Weise zu den HPC-Clustern und jedem anderen angeschlossenen Dienst, unabhängig vom verwendeten Zugang. Schließlich schafft die RegApp die Voraussetzungen für eine institutionenübergreifende Integration von Diensten und etabliert ein Gruppen-/

Rollenmanagement für überregionale und nationale Communities mit Delegationsmechanismen.

Die Software wird im Rahmen der Helmholtz-Plattform HIFIS³ und im aktuellen bwIDM-Projekt⁴ des Landes Baden-Württemberg eingesetzt und weiterentwickelt.

Impressum Datenschutz KIT

Übersicht Registrierte Dienste Dienste

Federated Login Service

Sie haben sich bereits bei den folgenden Diensten registriert:

bwSync&Share

bwSync&Share ist ein Online-Speicherdienst, der es ermöglicht, Ihre Daten zwischen verschiedenen Computern, mobilen Endgeräten und Benutzern zu synchronisieren bzw. auszutauschen und gleichzeitig in der Large Scale Data Facility (LSDF) am Karlsruher Institut für Technologie (KIT) zu sichern.

[Dienstbeschreibung](#)
[Registrierungsdetails](#)

bwDataArchive

Der Landesdienst bwDataArchive bietet Universitäten und öffentlichen Forschungseinrichtungen aus Baden-Württemberg eine technische Infrastruktur zur langfristigen Archivierung von Forschungsdaten. Er ermöglicht die Archivierung großer Datenbestände für einen Zeitraum von zehn oder mehr Jahren. Organisationen müssen einen Dienstleistungsvertrag mit dem KIT abschließen, um ihren Mitgliedern die Nutzung von bwDataArchive zu ermöglichen.

[Dienstbeschreibung](#)
[Registrierungsdetails](#)
[Dienstpasswort setzen](#)

bwUniCluster 2.0

Der am Steinbuch Centre for Computing (SCC) des Karlsruher Institut für Technologie (KIT) betriebene bwUniCluster 2.0 ist eines von mehreren zentralen Systemen für eine flächendeckende Grundversorgung der baden-württembergischen Universitäten und Hochschulen mit Hochleistungsrechnerkapazität.

[Dienstbeschreibung](#)
[Registrierungsdetails](#)
[Dienstpasswort setzen](#)
[SSH Key setzen](#)

Abbildung 3: RegApp-Registrierungsportal für Dienste

Innovative Two-Factor Authentication for secure HPC access

In spring 2020 cyberattacks on several HPC systems in Germany and Europe became known. They were based on successful logins with stolen credentials and SSH keys. SCC strengthened its HPC security by introducing a multifactor authentication scheme and decentralizing storage of SSH keys to a location outside of the HPC systems. The RegApp AAI software developed at SCC was extended to implement the necessary functions and interfaces. This implementation provides a significant overall security level increase in an innovative and future-proof manner, while making it possible to keep the necessary changes on the HPC system to a minimum.

More information about RegApp can be found on www.scc.kit.edu/dienste/regapp

³ hifis.net
⁴ bwidm.de

Performanzverbesserung des GridKa-Magnetbandspeichersystems

Ein umfangreicher Einsatz von Magnetbandspeichersystemen für spezielle Anwendungen und große Mengen an Forschungsdaten ist heute und auch in Zukunft sinnvoll. Am KIT existieren hierfür langjährige Erfahrung und die entsprechende Infrastruktur. Im Grid Computing Centre Karlsruhe (GridKa) konnte die Datenrate beim Lesen von Band durch umfangreiche Tests und die Entwicklung einer Spezialsoftware auf mehr als das Doppelte pro Laufwerk gesteigert werden. Damit wurde ein großer Fortschritt bei der Vorbereitung auf die kontinuierlich wachsenden Anforderungen der Experimente für die Hochenergie- und Astroteilchenphysik an das Magnetbandsystem erzielt.

Haykuhi Musheghyan, Andreas Petzold, Doris Rössmann

Magnetbandspeicher immer noch aktuell?

Die Entwicklung im Bereich der Massenspeichermedien wird aktuell einerseits durch Anwendungen wie Artificial Intelligence/Machine Learning (AI/ML) getrieben, die extreme Anforderungen an den Datendurchsatz stellen, andererseits steigt der Bedarf an Speicherplatz gerade in der Forschung immer weiter an. Für hohen Datendurchsatz werden verschiedene Flashspeicherlösungen verwendet, für große Speichervolumina kommen Festplatten zum Einsatz – oft auch in kombinierten Systemen. Magnetbandspeicher sind in den letzten Jahren an vielen Stellen durch andere Speicherlösungen verdrängt worden. Jedoch spielen sie in besonders großen Speicherinstallationen weiterhin eine wichtige Rolle und werden aktiv weiterentwickelt.

Ein Vorteil des Magnetbandspeichers sind die Betriebskosten. Seine Speichermedien sind im Verhältnis zum Festplattenspeicher günstig, wenn die Kosten pro Terabyte berechnet werden. Außerdem fallen Stromkosten nur bei aktiver Nutzung der Medien an. Der Betrieb von Festplatten verursacht dagegen dauerhaft Stromkosten. Bandbibliotheken, Laufwerke und Medien, die Hauptkomponenten eines Magnetbandspeichersystems, werden ständig weiterentwickelt. Dabei bietet besonders die Speicherdichte der Bandmedien noch mehrere Größenordnungen an Entwicklungspotential. Während im Festplatten-

bereich die jährliche Steigerungsrate der Speicherdichte auf ca. 15% gefallen ist, bestätigen die aktuellen Laborentwicklungen der Bandhersteller eine jährliche Steigerungsrate um ca. 30% [1]. Daten auf Band bieten auch eine zusätzliche Sicherheit in Bezug auf Cyberattacken. Die Daten auf einer Festplatte sind sehr schnell mit Schadsoftware „verseucht“ oder verschlüsselt. Es bleibt nur wenig Zeit, einen Angriff festzustellen und die entsprechenden Maßnahmen einzuleiten. Sind die Daten jedoch auf einem Bandmedium gespeichert, können diese nicht ohne weiteres geändert oder gelöscht werden. Selbst die auf einem Band gelöschten Daten können wiederhergestellt werden, solange sie nicht erneut überschrieben werden. Der wesentliche Nachteil des Magnetbandspeichers, nämlich der komplexe Zugriff auf die Daten mit hoher Latenz, stellt hier einen Vorteil dar. Selbstverständlich haben Bandmedien gerade durch ihre viel langsameren Zugriffszeiten auch deutliche Nachteile. Werden Daten regelmäßig bearbeitet und geändert, sollten diese nicht auf einem Bandmedium gespeichert werden. Für Archiv- und Backup-Zwecke, die die Hauptanwendungsfälle für Bandspeichersysteme darstellen, fällt dieser Nachteil nicht ins Gewicht.

Allerdings können mit auf die linearen Zugriffscharakteristika von Magnetband zugeschnittenen Zugriffsmustern auch Anwendungen unterstützt werden, die große Datenmengen zwar selten lesen, dafür

aber trotzdem hohe Lesegeschwindigkeiten benötigen. Bei optimaler Nutzung liefern Bandlaufwerke heute einen Datendurchsatz bis zu 400 MB/s. Die größten Bandbibliotheken am KIT enthalten jeweils über 30 Bandlaufwerke, so dass kumulative Datendurchsätze von mehreren GB/s realisierbar sind.

GridKa-Magnetbandspeichersystem

Der größte Nutzer von Magnetbandspeicher am KIT ist das Tier-1-Rechen- und Datenzentrum GridKa für die Experimente des Large Hadron Collider (LHC) am CERN in Genf. Bei GridKa werden ca. 15% der Rohdaten der vier LHC-Experimente auf Magnetband gespeichert (zurzeit etwa 60 PB). Außerdem nutzt auch das Belle-II-Experiment den GridKa-Magnetbandspeicher. Die Experimente lesen diese Daten regelmäßig in koordinierten Kampagnen, um neue Rekonstruktionssoftware und Kalibrationen anzuwenden.

Die erreichten Datenraten beim Lesen der Daten von Magnetband lagen mit nur 50 MB/s pro Bandlaufwerk über mehrere Jahre weit unterhalb der eigentlichen technischen Möglichkeiten, auch im Vergleich mit anderen Tier-1-Zentren. Deshalb wurde in 2018 ein Projekt ins Leben gerufen, das die Verbesserung des Datendurchsatzes des Magnetbandsystems mindestens auf 100 MB/s pro Bandlaufwerk zum Ziel hatte. Dazu wurde einerseits die gesamte Datentransferkette vom Bandlaufwerk bis zum festplattenbasierten Online-Speicher

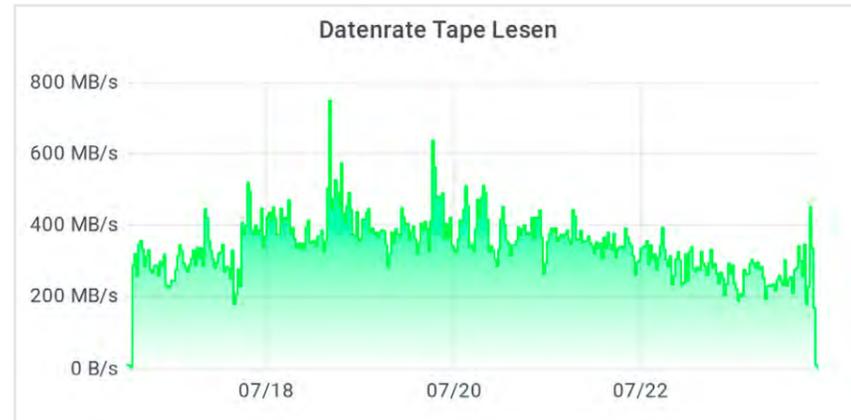


Abbildung 1: Leseratte von Band mit 8 Laufwerken vor der Optimierung. Die Leseratte pro Laufwerk beträgt im Durchschnitt etwa 40 MB/s.

von GridKa auf Engpässe untersucht und andererseits mit Benchmark-Datensätzen von etwa 190 TB Größe das Lesen der Daten von Band in verschiedenen Konstellationen getestet, um die im Produktivbetrieb genutzten Datenzugriffsmuster für einen möglichst hohen Datendurchsatz zu optimieren.

Perfomancetests und Verbesserungen

Ausgangspunkt war das bis dahin verwendete Softwaresetup, das jeweils maximal 2.000 Dateianfragen gleichzeitig an das Bandsystem übergeben konnte. Diese Anfragen werden von der am Andreas fragen entwickelten TSS-Software [2] nach Bändern und linearer Position auf Band sortiert und dann an das IBM Spectrum-Protect-Bandsystem [3] zur Abarbeitung übergeben. Der erreichte Datendurchsatz mit Parametern wie im Produktivbetrieb ist in Abbildung 1 dargestellt.

Bei der Analyse der Benchmark-Durchgänge wurden folgende Beobachtungen gemacht [5]:

Innerhalb der 2.000 Dateianfragen sind jeweils nur weniger als 100 Dateien von demselben Band, die wiederum auch nicht direkt nacheinander auf dem jeweiligen Band liegen. Das hat zur Folge, dass ein Band zwischen dem Lesen einzelner Dateien, lange gespult werden muss. Die Spulzeiten zwischen den Dateien betragen mehrere Minuten, da die eingesetzte Magnetbandtechnik zwar das ca. 1 km lange Band in zwei Minuten vom Anfang

bis zum Ende spulen kann, ein Band jedoch 144 Mal vom Anfang bis zum Ende beschrieben ist [4] und der Dateizugriff sequentiell erfolgt. Außerdem werden Bänder mehrfach eingelegt, um jeweils wenige Dateien zu lesen.

Um die Zeiten für das Spulen zu verringern, sollten also möglichst viele Dateien von einem Band gelesen werden, was wiederum bedeutet, dass dem Bandsystem von Anfang an deutlich mehr als 2.000 Anfragen zum Sortieren übergeben werden sollten. In einer separaten Testumgebung wurden daher Skalierungstests mit der TSS-Software durchgeführt, die eine maximal mögliche Anzahl von 30.000 gleichzeitigen Anfragen ergaben.

Abbildung 2 zeigt die in der Testumgebung erreichte Verdoppelung der Datenrate pro Laufwerk durch die dauerhafte

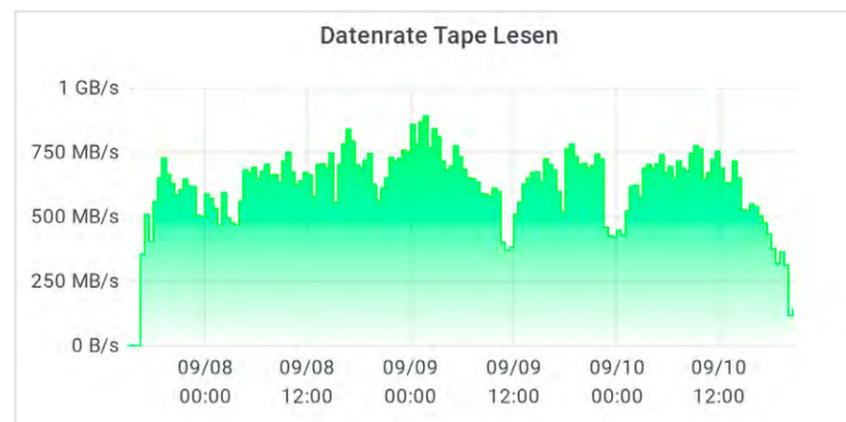


Abbildung 2: Leseratte von Band mit 30000 gleichzeitigen Anfragen und 8 Laufwerken. Die Leseratte pro Laufwerk beträgt im Durchschnitt etwa 80 MB/s.

Nutzung von 30.000 gleichzeitigen Dateianfragen an das Bandsystem. Der gelesene Datensatz ist dabei identisch zum Test in Abbildung 1. Diese Verbesserung demonstriert klar die Vorteile der großen Anzahl an Dateianfragen, die zu einer deutlichen Verringerung von Spulzeiten der Laufwerke führt.

Beim Durchführen der Benchmark-Tests wurde außerdem festgestellt, dass die Dateizugriffe sich auf viel mehr Bänder verteilen als erwartet. Als Grund konnte mehrfaches Schreiben einzelner Dateien auf verschiedene Bänder, verursacht durch zeitweise gestörtes Fehlermanagement, identifiziert werden, wobei immer die zuletzt geschriebene Kopie der Datei direkt lesbar ist. Diese Kopie wurde häufig mit großem zeitlichem Abstand zu anderen Dateien des ursprünglichen Datensatzes, und damit auf andere Bänder, geschrieben. Durch logisches Löschen der zusätzlichen Kopien konnten die optimal platzierten Dateien wieder lesbar gemacht werden und somit die Verteilung der Dateien über die Bänder wieder, wie ursprünglich erwartet, eingestellt werden.

In dem in Abbildung 3 dokumentierten Test konnte nach der Entfernung zusätzlicher Dateikopien erstmals eine Rate von mehr als 100 MB/s pro Laufwerk erreicht werden. Damit war in der Testumgebung das Ziel des Projekts Anfang 2019 erreicht.

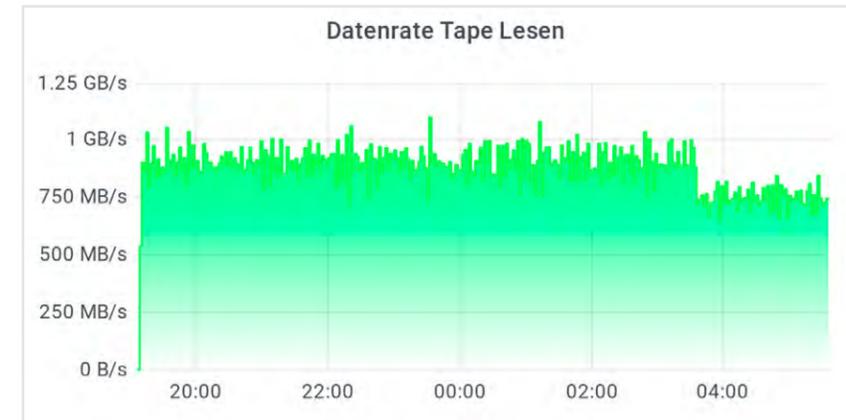


Abbildung 3: Leseratte von Band mit 8 Laufwerken nach Entfernung von zusätzlicher Dateikopien. Die Leseratte pro Laufwerk beträgt im Durchschnitt etwa 105 MB/s.

Umsetzung im Produktivbetrieb

Für die Umsetzung der erreichten Verbesserungen im Produktivbetrieb war die Entwicklung einer neuen Software notwendig, um die Limitierung der dort eingesetzten Software auf 2.000 gleichzeitige Dateianfragen zu überwinden. Basierend auf der am Tier-1-Zentrum NDGF [6] entwickelten ENDIT-Software [7], wurde bei GridKa die Software ENDIT-TSS entwickelt [8], die es ermöglicht, bis zu 30.000 gleichzeitige Dateianfragen so lange wie möglich im Bandsystem aktiv zu halten. Anfang 2020 konnte die ENDIT-TSS-Software in Betrieb genommen und die verbesserte Performance in Benchmark-Tests des ATLAS-Experiments, wie in Abbildung 4 gezeigt, erfolgreich demonstriert werden.

Erst durch den Einsatz von ENDIT-TSS war eine Nutzung von mehr als acht Laufwerken pro Experiment sinnvoll möglich. Dadurch konnte die lieferbare Datenrate von ca. 400 MB/s auf über 2 GB/s gesteigert werden. Inzwischen wurde ENDIT-TSS für alle Experimente bei GridKa eingeführt und ist erfolgreich im Produktiveinsatz.

Ausblick

Die bei der Optimierung und mit der Software ENDIT-TSS gewonnenen Erfahrungen fließen jetzt in die Migration aller GridKa-Daten von IBM Spectrum Protect nach HPSS [9] und in den zukünftigen Produktivbetrieb mit HPSS ein.

Performance improvement of the GridKa magnetic tape storage system

Extensive use of magnetic tape storage systems for special applications and large volumes of research data makes sense today and will continue to do so in the future. KIT has many years of experience in this field and the corresponding infrastructure. In the Grid Computing Centre Karlsruhe, the data rate for reading from tape could be increased by more than double per drive by sophisticated tests and the development of a special software. This was a major step forward in preparing for the continuously growing demands of experiments on the magnetic tape system.



Abbildung 4: Leseratte von Band mit 12 Laufwerken, optimiert mit ENDIT-TSS. Das Maximum von 2.2GB/s entspricht einer Leseratte von etwa 180MB/s pro Laufwerk.

Referenzen

- [1] Lantz, M. Why the Future of Data Storage is (Still) Magnetic Tape, IEEE Spectrum 2018, spectrum.ieee.org/computing/hardware/why-the-future-of-data-storage-is-still-magnetic-tape
- [2] TSS tsm-symposium.oucs.ox.ac.uk/2007/papers/Jos%20van%20Wezel%20-%20TSS.pdf
- [3] IBM Spectrum Protect (IBM SP), www.ibm.com/support/knowledgecenter/de/SSEQVQ/landing/welcome_sseqvq.html
- [4] Oracle T 10 Medien Spezifikation docs.oracle.com/cd/E48725_03/TKSA/appendix_a.htm
- [5] Musheghyan et al, The GridKa Tape Storage: various performance test results and current improvements, doi.org/10.1051/epjconf/20202450402_6
- [6] NDGF-T1, neic.no/nt1/
- [7] ENDIT github.com/neicnordic/endlit
- [8] Musheghyan et al, The GridKa tape storage: latest improvements and current production setup, to be published in proceedings of CHEP 2021, indico.cern.ch/event/948465/contributions/4323998/
- [9] High Performance Storage System (HPSS), hpss-collaboration.org

Spitzenplätze für HoreKa

Erst vor wenigen Wochen fertig aufgebaut und schon ganz vorne mit dabei: Der neue „Hochleistungsrechner Karlsruhe“ (kurz „HoreKa“) am SCC zählt aktuell zu den 15 schnellsten Rechnern in Europa. Bei der Energieeffizienz landet er weltweit sogar auf Platz 13.

Jennifer Buchmüller, Simon Raffener

In vielen Wissenschaftsbereichen sind Supercomputer nicht mehr aus dem Alltag von Forscherinnen und Forschern wegzudenken. Je schneller diese Rechner mathematische Gleichungen lösen und Daten verarbeiten, desto detailreicher und zuverlässiger werden die Simulationen, die damit betrachtet werden können. Um Wissenschaftlerinnen und Wissenschaftlern aus ganz Deutschland vor allem in den Erdsystemwissenschaften, den Materialwissenschaften, der Energie- und Mobilitätsforschung im Ingenieurwesen und der Teilchen- und Astroteilchenphysik ein Höchstmaß an Rechenleistung bieten zu können, hatte das KIT bereits Mitte 2019 mit der Beschaffung eines neuen Supercomputers begonnen. Der im Mai 2020 unterzeichnete Liefervertrag hatte eine Größenordnung von 15 Millionen Euro.

Herausgekommen ist ein innovatives Hybrid-System mit fast 60.000 Intel Xeon „Ice Lake“ Prozessorkernen, mehr als 220 Terabyte Hauptspeicher sowie 668 NVIDIA A100 Tensor Core GPUs. Als Kommunikationsnetzwerk kommt ein non-blocking NVIDIA Mellanox InfiniBand-HDR-Netzwerk mit 200 GBit/s pro Port zum Einsatz. Die Rechnersysteme stammen von Lenovo, die pro-com Datensysteme GmbH aus Eislingen bei Stuttgart übernahm die Projektkoordination und Lieferung.

Einer der 15 schnellsten Rechner Europas

Bereits während des Beschaffungsprozesses zeichnete sich ab, dass das System bei seiner Inbetriebnahme im Sommer 2021 voraussichtlich zu den schnellsten Rechnern Europas gehören würde. Maßgeblich für einen solchen Vergleich ist die „Top500“, die Liste der 500 schnellsten

Rechner der Welt. Nur die Systeme, die es auf diese Liste schaffen, dürfen sich „echte Supercomputer“ nennen.

Um einen Rechner auf der Liste platzieren zu können, muss die tatsächliche Rechenleistung mit einer speziellen Anwendung - dem sogenannten High Performance LINPACK Benchmark - vermessen werden. Schon während der Aufbauphase im März und April 2021 wurde daher mit den Benchmarks begonnen und HoreKa gemeinsam mit Intel, NVIDIA, Lenovo und ersten Nutzerinnen und Nutzern auf höchste Leistungsfähigkeit getrimmt.

HoreKa besteht aus zwei Komponenten: „HoreKa-Green“ umfasst die GPUs und „HoreKa-Blue“ die CPUs. Die Wertung von Hybridsystemen wie HoreKa ist durch die der Top500 zugrunde liegende Benchmark-Anwendung allerdings nicht vorgesehen. Das System taucht deswegen zwei Mal in der aktuellen Juni-Edition der Top500-Liste auf: einmal mit 8 PetaFLOPS auf Platz 53 und ein zweites Mal mit 2,33 PetaFLOPS auf Platz 220.

Ein PetaFLOPS entspricht dabei einer Leistung von einer Billion Rechenoperationen pro Sekunde. Insgesamt kann HoreKa sogar eine theoretische Spitzenleistung von 17 PetaFLOPS erbringen, was der Leistung von ca. 150.000 Laptops entspricht. Bei Anwendungen aus der Künstlichen Intelligenz sind sogar noch deutlich höhere Werte möglich.

Rechnen und Speichern gehen Hand in Hand

Ein zentraler Gesichtspunkt bei der Auslegung des Systems waren auch die enormen Datenmengen, die bei wis-

senschaftlichen Forschungsprojekten anfallen. Je nach Anwendung können von einer einzigen Simulation innerhalb von Stunden mehrere Hundert Terabyte an Daten erzeugt werden. Um mit den wachsenden Datenmengen Schritt zu halten, liefern die Rechenknoten, das InfiniBand-Netzwerk und die parallelen Dateisysteme von HoreKa im Vergleich zum Vorgängersystem ForHLR jeweils einen bis zu vier Mal höheren Speicherdurchsatz.

Als Datenablage dienen zwei parallele Spectrum-Scale-Dateisysteme mit einer Gesamtkapazität von mehr als 16 Petabyte. Eine mehrstufige Datenhaltung garantiert zusätzlich die Weiterverarbeitung auf externen Speichersystemen mit hohem Durchsatz. HoreKa ist beispielsweise mit bis zu 45 GByte/s Datenrate an den Online-Speicher der Large Scale Data Facility (LSDF) angebunden, die seit 2010 eine moderne Infrastruktur für die Speicherung, Verwaltung, Archivierung und Analyse von Forschungsdaten bietet und ebenfalls vom SCC betrieben wird.

Preisgekrönte Energieeffizienz

Supercomputer benötigen zwar viel Energie, diese wird aber sehr viel effizienter als bei herkömmlichen PCs und Laptops eingesetzt. HoreKa befindet sich in einem 2015 für den Vorgänger ForHLR neu errichteten Rechnergebäude auf dem Campus Nord des KIT. Das innovative Heißwasser-Kühlkonzept erlaubt es nicht nur, den Rechner ganzjährig mit minimalem Energieeinsatz zu kühlen, sondern im Winter können mit der Abwärme auch die Büroräume beheizt werden. Dieses im Jahr 2017 mit dem Deutschen Rechenzentrumspreis ausgezeichnete Konzept



Abbildung 1: Der neue Supercomputer HoreKa in den Rechnerräumen des SCC am Campus Nord (Foto: Simon Raffener/SCC)

wurde beim neuen System beibehalten und konsequent weiterentwickelt.

Kein Wunder also, dass HoreKa auch in puncto Energieeffizienz Rekorde bricht. Aktuell steht das System auf Platz 13 der „Green500“, der Liste der energieeffizientesten Supercomputer weltweit. Das KIT nimmt mit seiner Größe und technischen Ausrichtung bei der Realisierung einer nachhaltigen Zukunft eine herausragende Stellung ein - nicht nur in der Wissenschaft, sondern auch beim effizien-

ten Betrieb der Forschungsinfrastruktur. Hier ist das KIT jetzt schon international mit führend. Um weiterhin an der Spitze zu bleiben, wird dieser Bereich auch in Zukunft weiter ausgebaut, beispielsweise werden am SCC weitere Möglichkeiten der Wärmenachnutzung evaluiert.

Offizielle Einweihung im Juli 2021

Der neue Supercomputer ist aber nicht nur ein großer Erfolg für den Standort Karlsruhe, sondern auch für das Nationale

Hochleistungsrechnen (NHR) in Deutschland. Mit HoreKa bringt das KIT nur ein halbes Jahr nach Gründung des NHR-Verbundes als erstes von insgesamt acht NHR-Zentren einen neuen Rechner in die Allianz ein. Die offizielle Einweihungsfeier mit der Übergabe an die Wissenschaft fand am 30.07.2021 statt.

Mehr Informationen zum neuen NHR-Zentrum und HoreKa finden Sie unter www.scc.kit.edu/dienste/nhr

New Supercomputer at KIT scores top positions

SCC has procured a new Tier 2 HPC cluster system including a new data storage system. The system is called "Hochleistungsrechner Karlsruhe", short "HoreKa". It is currently ranked 53rd on the list of the fastest Supercomputers (Top500) and 13th on the list of the most energy-efficient Supercomputers (Green500) in the world. This makes HoreKa one of the 15 fastest Supercomputers in Europe.

The new system has more than 60.000 Intel Xeon processor cores, more than 220 Terabytes of RAM as well as 668 NVIDIA A100 Tensor Core GPUs. A non-blocking InfiniBand HDR network with 200 GBit/s per port is used for communication between the nodes. Two Spectrum Scale parallel file systems offer a total storage capacity of more than 15 petabytes. HoreKa is also connected to the "Large Scale Data Facility" (LSDF) with a data rate of up to 45 GByte/s.

The system has been installed in a state-of-the-art data center constructed for its predecessor ForHLR on KIT's Campus North in 2015.

More information about HoreKa and the NHR center can be found under www.scc.kit.edu/en/services/nhr.

KONKIT – ein Beitrag zur Kontrolle der Pandemie

Trotz der zahlreichen virtuellen Formate erfordert der Lehrbetrieb am KIT, dass Menschen zusammenkommen. Gemäß den Vorgaben der Corona-Verordnung ist das KIT verpflichtet, in diesem Fall die Kontaktnachverfolgung zu gewährleisten. Um dabei den Aufwand vor Ort gering zu halten und gleichzeitig die Gesundheitsbehörden bestmöglich bei der Nachverfolgung von Kontakten infizierter Personen zu unterstützen, haben die zuständigen Einrichtungen des KIT gemeinsam mit dem SCC eine in die IT-Infrastruktur integrierte Lösung entwickelt.

Axel Maurer, Patrick von der Hagen

Gemeinsames Vorgehen und gegenseitige Verlässlichkeit ist ein Schlüssel zur Kontrolle der Pandemie. Nach diesem Grundsatz handelt das KIT auch im Schulterschluss mit den Gesundheitsbehörden, wenn es um das wichtige Instrument der Kontaktnachverfolgung bei Zusammenkünften geht. Trotz Virtualisierung und Digitalisierung der Lehre lassen sich nicht alle Begegnungen vermeiden, sei es bei der praktischen Ausbildung der Studierenden im Rahmen von Laborpraktika oder bei einem Teil der Prüfungen.

Neben funktionierenden Hygienekonzepten ist es erforderlich, die Teilnahme an Veranstaltungen zu dokumentieren, um die Gesundheitsbehörden im Bedarfsfall mit akkuraten Teilnehmerlisten unterstützen zu können.

Auch bei einer sukzessiven Wiederaufnahme des Präsenzbetriebs ist zu erwarten, dass eine mögliche Kontaktnachverfolgung eine wichtige begleitende Maßnahme darstellen wird. Daher wurde eine Arbeitsgruppe etabliert, um ein Gesamtkonzept zur Kontaktnachverfolgung am KIT (KONKIT) zu entwerfen. Beteiligt an dieser Arbeitsgruppe waren Vertreterinnen und Vertreter der Fakultäten, der Institute, der Dienstleistungseinheiten „Studium und Lehre“ und „Allgemeine Services“, des SCC sowie des Personalrats und des Datenschutzteams.

Dabei sollte eine Lösung erarbeitet werden, die

- eine Kontaktnachverfolgung gemäß Infektionsschutzgesetz (IfSG) gewährleistet

- den Erfordernissen des Datenschutzes entspricht
- eine zentrale Bearbeitung von Anfragen durch einen eingeschränkten Mitarbeiterkreis ermöglicht
- für mehrere hundert Räume, vom kleinen Labor bis zum Vorlesungszelt, skaliert
- einen hohen Durchsatz bei der Erfassung gewährleistet, um Gedränge zu vermeiden
- robust, wartungsarm und ausfallsicher ist
- sowohl von Mitarbeitenden und Studierenden als auch von Gästen genutzt werden kann

Mit diesen Zielen wurde eine hybride Lösung mit Registrierung per Smartphone auf Basis eines QR-Codes oder stationär mit der KIT-Card entwickelt. Für die Teilnehmenden sind zwei Säulen sichtbar. Zum einen die „KONKIT-Box“, die den persönlichen KIT-Ausweis auslesen kann und insbesondere in den großen Hörsälen zum Einsatz kommt, zum anderen die KONKIT-QR-Codes, die sowohl in den Hörsälen als auch in kleinen Laboren ausgehängt und genutzt werden können. Beschäftigte und Studierende haben

dann die Wahl, entweder die Kartennummer über eine KONKIT-Box oder ihr KIT-Konto über den QR-Code zu erfassen. Die Datenverarbeitung erfolgt dann unter Pseudonym, sofern nicht tatsächlich eine Anfrage gemäß IfSG eingeht, die eine Auflösung des Pseudonyms erfordert. Gäste müssen allerdings über den QR-Code in jedem Fall einen vollständigen Datensatz erfassen.

Bei der Umsetzung des Konzeptes galt es, folgende Punkte zu berücksichtigen:

1. Teilnehmende an einer Veranstaltung registrieren ihre Anwesenheit entweder kontaktlos über den persönlichen KIT-Ausweis und ein Kartenlesegerät oder über einen aushängenden QR-Code.
2. Die betreffenden Personen haben jederzeit die Möglichkeit, sich über die verarbeiteten Daten und die Form, wie sie verarbeitet werden, zu informieren. Sollten die technischen Lösungen nicht genutzt werden, muss es eine Alternative zur digitalen Kontaktnachverfolgung geben.
3. Angehörige des KIT werden pseudonym erfasst (verwendete Karte, persönliche Kontobezeichnung).

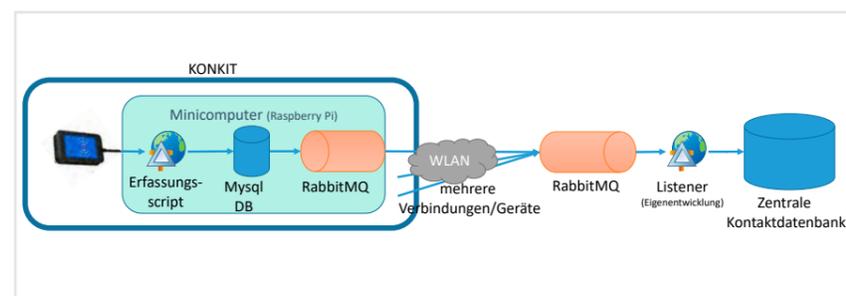


Abbildung 1: Schema der KONKIT-Struktur

Diese Daten werden nur im Falle einer tatsächlichen Anfrage durch die Gesundheitsbehörden mit den bereits vorhandenen Stammdaten zusammengeführt, um eine Liste der Kontakte zu erzeugen.

4. Prinzipiell wird beim KONKIT-System die Zusammenkunft in einem Raum dokumentiert. Dazu wird der Zeitpunkt des Eintritts und des Verlassens des Raumes erhoben. Bei Veranstaltungen, bei denen das Ende feststeht, wie beispielsweise Lehrveranstaltungen, übernimmt das System selbständig die Buchung des Endes der Zusammenkunft, so dass die Teilnehmenden sich nur zu Beginn der jeweiligen Veranstaltung einbuchen müssen.

Die KONKIT-Box

Für die schnelle und zuverlässige Erfassung der Teilnehmenden per KIT-Card hat das SCC die KONKIT-Box konzipiert. Sie ermöglicht es, sich mit der auf dem Chip der KIT-Card gespeicherten bwCard-Nummer schnell und einfach für eine Zusammenkunft zu registrieren. Technisch wird für die KONKIT-Box ein Minicomputer (Raspberry Pi) und ein speziell programmierter Chipkartenleser (Elatec) genutzt. Der Kartenleser liest die bwCard-Nummer von der KIT-Card, verschlüsselt diese und übergibt die verschlüsselte Kartennummer an den Minicomputer. Es gibt also keine Möglichkeit, anhand der Daten auf dem Minicomputer ein Anwesen-



Abbildung 2a und 2b: Gehäusevarianten der KONKIT-Box

heitsprofil einzelner Personen zu erstellen. Dort wird die verschlüsselte bwCard-Nummer zusammen mit der Uhrzeit und der Raum-Kennung in einer Datenbank zwischengespeichert und asynchron über ein Message-Queue-System an die zentrale Datenbank gesendet. Die Daten auf dem Minicomputer werden 24 Stunden nach der Übergabe an das Zentralsystem gelöscht. Abbildung 1 zeigt die KONKIT-Struktur. In Abbildung 2a und 2b sieht man unterschiedliche Gehäusevarianten der KONKIT-Box.

KONKIT-QR

Als zweite Säule des Erfassungssystems von KONKIT gibt es die Möglichkeit, für Räume, die im Campus-Management-System des KIT erfasst sind, Poster mit Erläuterungen und QR-Code aufzuhängen. Dieser QR-Code enthält eine URL, mit der man sich am KONKIT-System für einen Raum ein- und wieder ausbuchen kann. Dabei besteht die Möglichkeit, entweder den KIT-Account oder (insbesondere für externe Teilnehmerinnen und Teilnehmer) eigene Kontaktdaten zu hinterlegen. Die Anmeldedaten bleiben mit Zustimmung der Nutzenden auf dem Handy zur Wiederverwendung erhalten, um künftige Anmeldungen zu vereinfachen.

Datenhaltung

Zusätzlich zur bwCard Nummer bzw. den bei KONKIT-QR erhobenen Daten werden noch der Zeitpunkt der Buchung und die Raumkennung in einer zentralen Datenbank erfasst. Neben den Raumdaten sind dort auch die Veranstaltungsdaten hinterlegt, so dass ein automatisches Check-Out erfolgen kann. Die personenbezogenen Daten werden nach der vorgegebenen Aufbewahrungsfrist aus der Datenbank gelöscht.

Umgang mit Anfragen

Sollten die Gesundheitsbehörden eine Anfrage zur Kontaktermittlung stellen, so wird diese immer an die zentral zuständigen Mitarbeitenden weitergeleitet. Diese

prüfen die Anfrage, halten ggf. Rücksprache mit den Dozierenden und haben die Möglichkeit, die Kontaktdaten über eine speziell entwickelte Web-Anwendung zu erstellen und die Pseudonyme aufzulösen. Die Daten werden dann entsprechend den Vorgaben der Gesundheitsbehörden an diese übermittelt. Insgesamt sind derzeit rund 80 KONKIT-Boxen in den größeren Lehrveranstaltungsräumen aufgestellt. Für über 630 Räume wurden QR-Codes hinterlegt. Seit Februar wurden über das KONKIT-System über 30.000 Kontakte registriert.

Fazit

Mit dem KONKIT-System, bestehend aus KONKIT-Box und KONKIT-QR, hat sich das KIT bestens auf die erforderliche Kontaktnachverfolgung im Infektionsfall vorbereitet. Das KONKIT-System hat sich bewährt und erfährt eine hohe Akzeptanz, da es sich durch geringen Aufwand für die Beteiligten und Betroffenen auszeichnet und die Gesundheitsbehörden mit validen Daten bei der Kontaktnachverfolgung unterstützen kann. Auch bezüglich einer eventuellen Erweiterung im Falle einer erforderlichen Überprüfung, ob eine Person genesen oder geimpft ist, gibt es für das KONKIT-System bereits Überlegungen – natürlich immer in der Hoffnung, dass das System so wenig wie möglich zum Einsatz kommen muss.

KONKIT – a Contribution towards Controlling the Pandemic

Despite the numerous virtual formats, teaching at KIT requires that people come together. According to the requirements of the Corona Regulation, KIT is obliged to provide contact tracing in this case. To keep the on site effort low and at the same time to support the health authorities in the best possible way in tracing contacts of infected persons, the responsible KIT institutions have developed a solution integrated into the KIT infrastructure together with SCC. With the KONKIT system deployed since October 2020, KIT thus makes an important contribution to limiting the spread of COVID-19.

WLAN für IoT-Geräte am KIT

In der Vergangenheit war es am KIT nicht möglich, IoT-Geräte, die keine Authentifizierung per WPA2-Enterprise unterstützen, mit dem WLAN zu verbinden. Dies wird nun durch das im Juli 2021 eingeführte WLAN "KIT-IoT" am KIT mittels der Technik Multi Pre-Shared Key (MPSK) ermöglicht.

Janis Streib, Klara Mall

Das Internet of Things (IoT) beschreibt die Realität einer immer weiter wachsenden Vernetzung von Geräten. Auch am KIT gibt es immer mehr Geräte mit einer WLAN-Schnittstelle, die Netzkonnektivität erhalten sollen. Ob verteilte Sensoren, eingebettete Systeme oder Streaming-Media-Adapter, am KIT gibt es dafür vielfältige Anwendungsfälle.

Bisher eingesetzte Verfahren und Anforderungen

IoT-Geräte unterstützen in der Regel nicht das Authentifizierungsverfahren IEEE 802.1X, was beim sogenannten WPA2-Enterprise zum Einsatz kommt. Diese Technik wird auch in den SSIDs KIT und eduroam am KIT genutzt. Sie erlaubt die Authentifizierung mit Nutzernamen und Passwort, was bei großen Umgebungen mit vielen Organisationseinheiten, Nutzern und Clients vorzuziehen ist, da nicht ein gemeinsames Passwort unter allen Nutzern geteilt werden muss. Dies kennt man aus dem heimischen WLAN, in dem meist WPA2-PSK zur Authentifizierung zum Einsatz kommt, bei dem eine gemeinsame Passphrase (Pre-Shared Key; PSK) allen Nutzenden bekannt ist.

Die Nutzung eines gemeinsamen Schlüssels ist kein gangbares Verfahren in einer so großen Einrichtung wie dem KIT. Hier benötigt jede Person oder eben jedes Gerät ein eigenes Passwort für das WLAN. In der Vergangenheit gab es somit keine Möglichkeit, den IoT-Geräten, die nur WPA2-PSK unterstützen, WLAN-Konnektivität zu geben.

Darüber hinaus ist eine Kommunikation mit IoT-Geräten oder zwischen diesen erforderlich, was im allgemeinen Netz für die SSIDs KIT und eduroam aus Sicher-

heitsgründen prinzipiell verboten ist. Somit scheidet es aus, alle IoT-Geräte in ein gemeinsames Netz zu setzen. Vielmehr sollen beispielsweise nur die Sensoren innerhalb eines verteilten Systems oder ein Instituts-Laptop und ein Streaming-Media-Adapter miteinander kommunizieren. Daher müssen dedizierte Netze für die Anbindung von IoT-Geräten zur Verfügung stehen.

MPSK macht es möglich

Mit der Einführung von Multi Pre-Shared Key (MPSK) hat der Hersteller Aruba, dessen Komponenten wir im WLAN am KIT einsetzen, eine Erweiterung zu WPA2-PSK geschaffen, die die Nutzung eines gerätespezifischen Keys zur Authentifizierung erlaubt. Somit kann für jedes Gerät ein eigener Key genutzt werden, ohne dass sich auf der Client-Seite des IoT-Geräts die Technik vom herkömmlichen WPA2-PSK-Standard unterscheidet. Hierbei wird das Gerät anhand seiner MAC-Adresse

identifiziert und diesem ein Pre-Shared Key zugewiesen (Abbildung 1).

Seit Juli 2021 bieten wir am KIT diese neue Methode unter der SSID KIT-IoT an. Über das Netzdienste-Verwaltungssystem NETVS¹ können IT-Beauftragte nun IoT-Clients über den Menüpunkt MACAuth in ihren Netzen anlegen. Über die Option *WPA-Schlüssel (neu) generieren* wird beim Eintragen eines Clients ein Zugangsschlüssel erzeugt, der dann für das IoT-Gerät verwendet werden kann. Die Grundfunktionalität im NETVS steht bereits seit November 2020 zur Verfügung – jedoch bislang nur zur Verwendung mit kabelgebundenen Clients.

Technische Umsetzung

Um diese Technik zur Verfügung stellen zu können, mussten zunächst einige Vorarbeit geleistet werden. Authentifiziert sich ein Client mit dieser Methode, reicht der WLAN-Controller die MAC-Adresse

Client anlegen
✕

BCD-Name

Name der Broadcastdomain

Client-MAC-Adresse

MAC-Adresse zur Client-Identifikation

MACAuth-Client-Beschreibung

MACAuth-Client-Info

WPA-Schlüssel (neu) generieren (nur für WiFi-Clients)

Bei create: Neuen Client-WPA-Schlüssel erzeugen. Bei update: Schlüssel erneuern. Wird nur für WiFi-Clients benötigt.

Abbrechen
Anlegen

Abbildung 1: Anlegen eines neuen MPSK-Clients im NETVS

des Clients an den Authentifizierungs-Server (RADIUS-Server) weiter. Dieser prüft, ob die MAC-Adresse für KIT-IoT zugelassen ist und wenn ja, welcher Pre-Shared Key und welches VLAN für diese hinterlegt sind. Wenn sie nicht zugelassen ist, sendet er Access-Reject an den WLAN-Controller zurück, und der Client wird abgewiesen. Im anderen Fall werden Access-Accept und der zugehörige Key sowie die VLAN-ID an den WLAN-Controller zurückgegeben. Der WLAN-Controller sendet den Key an den Access Point, über den sich der Client mit dem WLAN verbinden möchte, und dieser kann dann – sofern der Key mit dem vom Client gesendeten übereinstimmt – die Verbindung herstellen (Abbildung 2).

Der Controller weist hierbei dem Client das VLAN zu, das der RADIUS-Server für ihn zurückgegeben hat. Da bei der VLAN-Zuweisung für den Client dieselbe Technik wie bei WiFi2VLAN² zum Einsatz kommt, muss die Aktivierung von MPSK für das entsprechende Netz beantragt werden.³ Im Vergleich zu WiFi2VLAN unterscheidet sich lediglich die Authentifizierung. In einen Fall handelt es sich um Nutzerauthentifizierung mit dem KIT-Account und 802.1X, im anderen Fall mit der MAC-Adresse und MPSK.

Das Authentifizierungs-Backend besteht aus einem Freeradius-Server und einem Python-Skript, das unter der Verwendung des Freeradius-Python2-Moduls⁴ eingebunden wird. Hierbei kommt noch Python 2 zum Einsatz, da das Freeradius-Python3-Modul nach Angaben des Projekts zum aktuellen Zeitpunkt noch nicht stabil ist. Das Skript lässt sich jedoch ohne großen Aufwand im Nachhinein auf Python 3 portieren.

Das Python-Modul dient zur Interaktion mit der NETDB – der Postgresql-Datenbank, die unter anderem hinter dem NETVS arbeitet und die Daten zur MPSK-Authentifizierung und die Zuordnung des Clients (identifiziert anhand seiner MAC-Adresse) zur Broadcast-Domain bereitstellt. Aufgrund der Identifizierung des Clients anhand seiner MAC-Adresse kann eine MAC-Adresse nur genau einmal im

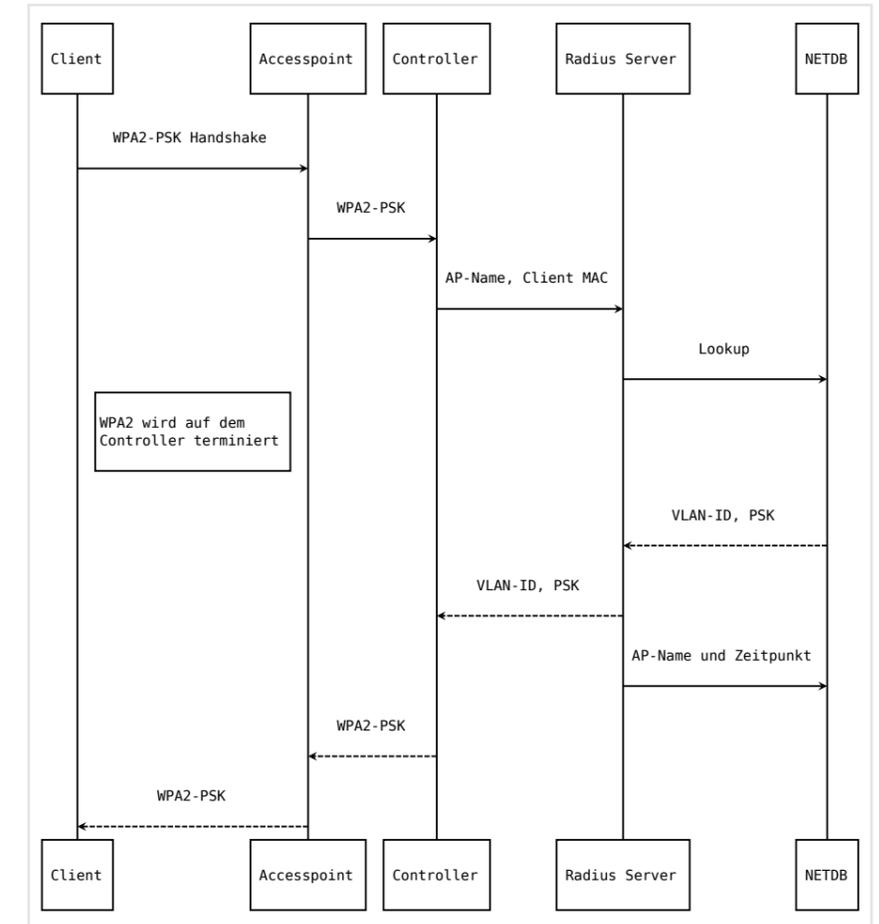


Abbildung 2: Schematische Darstellung einer erfolgreichen Anmeldung im WLAN mittels MPSK

gesamten KIT eingetragen werden. Die Datenbankabfragen zur Authentifizierung werden zukünftig in einer lokalen Read-Only-Replikation der Datenbank stattfinden, um eine höhere Ausfallsicherheit des Dienstes zu erzielen.

Bei erfolgreicher Anmeldung wird in der Datenbank – und somit auch im NETVS sowie in der NET-API – der Access Point sowie der Zeitpunkt der letzten Anmeldung vermerkt. Dies erlaubt eine bessere Kontrolle über die Benutzung und Aktivität der Clients. Das Python-Skript sowie eine Freeradius-Beispielkonfiguration⁵ stehen öffentlich unter freier Lizenz zur Verfügung.

WLAN for IoT Devices at KIT

In the past, it was not possible at KIT to connect IoT devices that do not support authentication via WPA2-Enterprise to the WLAN. Now, this is made possible by the KIT-IoT-WLAN introduced in July 2021 using Multi Pre-Shared Key (MPSK) technology.

¹ netvs.scc.kit.edu/macauth

² Direkter WLAN Zugang in institutseigene Netze (www.scc.kit.edu/dienste/wifi2vlan.php)

³ www.scc.kit.edu/dienste/internet-of-things.php

⁴ wiki.freeradius.org/modules/Rlm_python

⁵ git.scc.kit.edu/scc-net/nac.py

Studenten-Meldeverfahren der gesetzlichen Krankenversicherungen

Das KIT nimmt als Pilotkunde seit dem 1. April 2021 am Studenten-Meldeverfahren (SMV) der gesetzlichen Krankenversicherungen zum elektronischen Austausch von Versicherungsnachweisen von Studierenden teil. Hierdurch entfällt der Nachweis des Versichertenstatus durch Studierende bei der Immatrikulation in Papierform. Das elektronische Meldeverfahren wird ab 2022 per Gesetz für alle Hochschulen in Deutschland verpflichtend.

Jana Trummer-Wilke, Philip Hoyer

Der Datenaustausch zwischen den gesetzlichen Krankenkassen (GKV) und den Arbeitgebern erfolgt für gesetzlich versicherte Angestellte bereits seit vielen Jahren elektronisch. Die Meldung der Hochschulen an die GKV erfolgte dagegen aufgrund einer Verordnung immer noch in Papierform. Bei der Immatrikulation mussten Studierende ihren Krankenversicherungsstatus nachweisen und die Hochschulen haben diese Informationen an die Krankenkassen übermittelt – beides über den Postweg. Der Gesetzgeber hat 2019 beschlossen, dass das sogenannte Studenten-Meldeverfahren (SMV) zukünftig elektronisch durchzuführen ist und dies per Gesetz geregelt (§199a SGB V). Mit der konkreten Ausgestaltung wurden die Hochschulrektorenkonferenz (HRK) und der GKV-Spitzenverband betraut. Das SMV ist an das bereits bestehende Angestellten-Meldeverfahren angelehnt.

Bereits seit dem 01.01.2021 sind die gesetzlichen Krankenkassen verpflichtet, die Meldungen zum Status ihrer versicherten Studierenden elektronisch an die Hochschulen zu übermitteln. Voraussetzung ist, dass die Hochschulen sich bereit erklärt haben, am elektronischen Verfahren teilzunehmen. Erst ab dem 01.01.2022 ist die Teilnahme an dem SMV für die Hochschulen verpflichtend.

Das KIT hat sich in Zusammenarbeit mit der HIS eG, deren Campus-Management-Software am KIT für das Studierendenmanagement eingesetzt wird, bereiterklärt, als Pilothochschule zu fungieren und die noch laufende Entwicklung in den HIS-Modulen SOS-GX bzw. BSOS zu unterstützen. Dies bedeutet, dass das KIT seit dem 01.04.2021 am elektronischen Datenaustausch mit den Krankenkassen teilnimmt.

Bewerberinnen und Bewerber teilen die Absicht, ein Studium zu beginnen, ihrer Krankenkasse z.B. über eine App oder ein Online-Portal mit. Die Krankenkasse übermittelt den Versichertenstatus auf elektronischem Weg an die Hochschule. Von der gesetzlichen Versicherungspflicht befreite Personen müssen dies bei einer gesetzlichen Krankenkasse ihrer Wahl nachweisen, damit der Befreiungstatbestand ebenfalls als elektronische Bestätigung an die Hochschule übermittelt werden kann. Bei der Immatrikulation im Studierendenservice wird automatisch der Versichertenstatus der Bewerbenden geprüft. Sofern erfolgreich, wird der Krankenkasse auf gleichem Weg der Beginn des Studiums mit Tag der Einschreibung und Semesterstart, gemeldet – ebenso bei Exmatrikulation das Ende des Studiums mit Tag der Exmatrikulation und des Semesterendes.

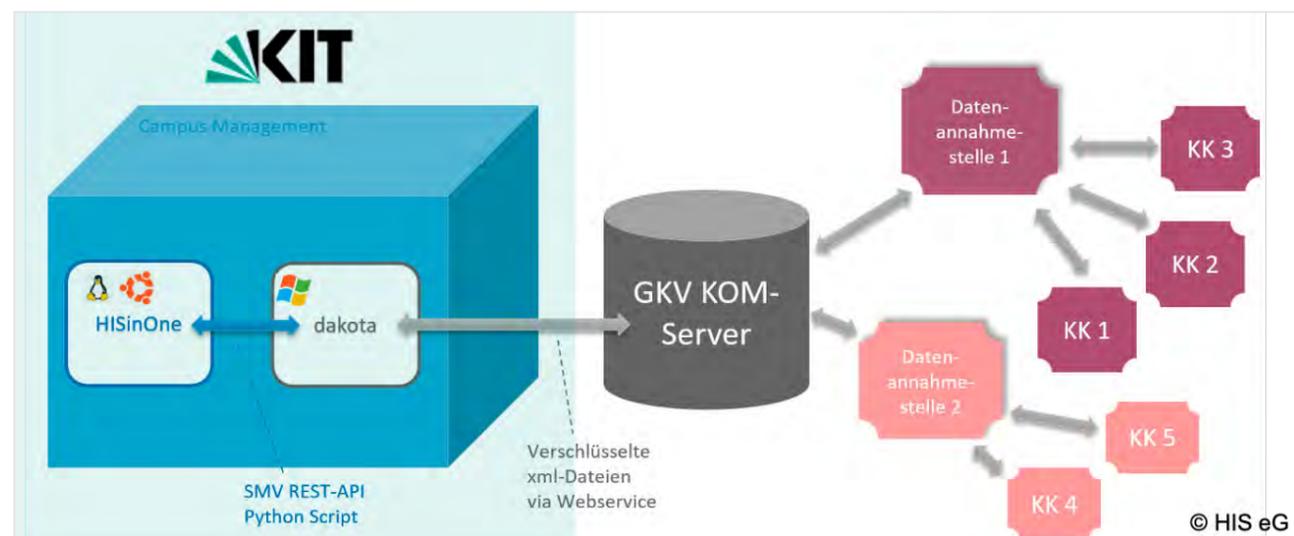


Abbildung 1: Am SMV beteiligte Softwaresysteme und deren Kommunikationsschnittstellen

Weitere Ereignisse, die einen Datenaustausch zwischen Hochschulen und Krankenkassen auslösen können, sind z.B. der Verzug von Beitragszahlungen (dies löst eine Rückmeldesperre aus) oder ein Wechsel der Krankenkasse.

Die Vorarbeiten für die Inbetriebnahme der Software-Module starteten bereits Ende 2020. Dazu gehörten unter anderem die Beantragung einer speziellen Absennummer und eines Zertifikates für das KIT zur Nutzung der im Verfahren eingesetzten Software dakota (Datenaustausch und Kommunikationen auf der Basis Technischer Anlagen). Dakota ist ein Programm der Informationstechnischen Servicestelle der gesetzlichen Krankenversicherung (ITSG) zum gesicherten Übertragen von Sozialversicherungsmeldungen. Der Datenaustausch zwischen Hochschulen und Krankenkassen wird über den zentra-

len GKV-Kommunikationsserver abgewickelt. Die am KIT eingesetzte HIS-Software erzeugt und liest XML-Dokumente, die von dakota ver- bzw. entschlüsselt und per Webservice zum GKV-Server gesendet bzw. abgeholt werden. Die Krankenkassen senden und empfangen über verschiedene Datenannahmestellen die entsprechenden Meldungen vom GKV-Server (Abbildung 1).

Die Zusammenarbeit mit der HIS eG gestaltete sich angenehm und war geprägt von kurzen Reaktionszeiten und effektiven Besprechungen zur Behebung etwaiger Probleme. Auch wurden immer wieder das Feedback des KIT zum aktuellen Stand der Softwareentwicklung eingeholt und Verbesserungsvorschläge, um beispielsweise Prozessschritte besser zu unterstützen, dankend angenommen und rege diskutiert. Das SCC und die Dienst-

leistungseinheit Studium und Lehre (SLE) blicken nun gespannt auf die nächste Bewerbungsphase, in der sich das SMV im Praxistest beweisen muss.

Weitere Informationen zu dem Studenten-Meldeverfahren sind auf den Webseiten von SLE verfügbar:

<https://www.sle.kit.edu/vorstudium/krankenversicherung.php>

Student Reporting Procedure of the statutory health insurances

KIT has been participating as a pilot customer in the Student Reporting Procedure (SMV) of the statutory health insurance companies for the electronic exchange of student insurance certificates since 1st April 2021. This eliminates the need for students to provide proof of insurance status in paper form when enrolling. The electronic reporting procedure will become mandatory by law for all universities in Germany from 2022.

EPICUR Inter-University Campus

Mit der Virtual Campus Learning Platform (VCLP) hat das SCC eine wichtige Aufgabe in der Lehre in der EPICUR-Allianz gelöst. Die VCLP stellt einen zentralen Zugangspunkt zum gesamten EPICUR-Lehrangebot bereit und integriert die bestehenden Lern-Management-Systeme an den acht Partneruniversitäten. Studierende und Lehrende können sich einfach über ihre bestehenden Uni-Accounts anmelden und an den angebotenen EPICUR-Kursen auf den verschiedenen Lehr- und Lernplattformen direkt teilnehmen.

Philip Hoyer

Das KIT ist Mitglied der EPICUR-Allianz (European Partnership for an Innovative Campus Unifying Regions), ein Verbund aus acht europäischen Universitäten. Neben dem KIT sind die Universitäten aus Freiburg, Straßburg, Mulhouse, Amsterdam, Thessaloniki, Poznan, sowie die Universität für Bodenkunde aus Wien Mitglieder der Allianz (Abbildung 1). Ziel der Allianz ist es unter anderem, die Mobilität von Studierenden zu fördern und gemeinsame Kursangebote zu etablieren, die insbesondere in der Pandemie auch virtuell, d.h. ohne physische Präsenz, besucht werden können.

EPICUR Inter-University Campus (EIUC)

Zur Erreichung der Ziele sind die Arbeiten innerhalb der EPICUR-Allianz in verschiedene Work Packages gegliedert. Im Work Package 4 sollen die entsprechenden IT-Plattformen und -Infrastrukturen aufgebaut werden, um Studierende und Lehrende in EPICUR zu unterstützen. Trotz der erschwerten Pandemiebedingungen konnten die beteiligten Mitarbeiterinnen und Mitarbeiter den sogenannten EPICUR Inter-University Campus (EIUC) in sehr kurzer Zeit aufbauen und das EPICUR-Lehrangebot über diesen abwickeln.

Der EIUC besteht im Wesentlichen aus zwei miteinander integrierten Systemen:

- Die Virtual Campus Learning Platform (VCLP) ist das zentrale EPICUR Lern-Management-System, welches vom SCC entwickelt und betrieben wird. Das SCC hat hierbei zusammen mit der Universität Freiburg auf die Erfahrungen mit der Lernplattform ILIAS zurückgegriffen und eine neue ILIAS-Plattform unter der Web-Adresse learn.epicur.education aufgebaut. Gleichzeitig war es von Anfang an ein Ziel, auch die lokalen Lern-Management-Systeme an



Abbildung 1: Die acht Universitäten der EPICUR-Allianz.

den einzelnen Partneruniversitäten mit einzubeziehen und diese in die zentrale EPICUR-Plattform zu integrieren.

- Das Virtual University Registrar System (VURS) ist das zentrale Campus-Management-System (bzw. System Information System), das von den Partnern der Aristotle Universität Thessaloniki (AUTH) entwickelt wurde. Die Universität hat auf das von ihr und anderen griechischen Universitäten gemeinsam entwickelte Campus-Management-System Universis zurückgegriffen und eine für EPICUR angepasste Version entwickelt.

Beide Systeme – VCLP und VURS – sind miteinander integriert und tauschen über die vorhandenen Schnittstellen Daten aus. So werden z.B. Kursanmeldungen aus VURS an VCLP übertragen. Umgekehrt werden die Kursinformationen, die auf der VCLP angezeigt werden, direkt aus dem VURS bezogen, um die Pflege der Kursinformationen auf eine Stelle zu beschränken und den Aufwand für die Kursanbieter zu minimieren.

Anmelden mit dem Account der Heimat-Universität

Da alle an EPICUR beteiligten Universitäten auch an der eduGAIN-Föderation teilnehmen, war es, nach der aufwändigen Klärung der datenschutzrechtlichen Aspekte, technisch relativ einfach, allen Partnern den Zugang über die bestehenden Heimat-Accounts zu ermöglichen. Mittlerweile sind sowohl VCLP, als auch VURS über Shibboleth angebunden. Mitarbeitende und Studierende wählen bei der Anmeldung zunächst ihre Heimateinrichtung aus und melden sich dann am lokalen Identity Provider an. Grundlegende Stammdaten, wie Vor- und Nachname, E-Mail-Adresse und der eindeutige Accountname (EduPersonPrincipalName, EPPN) werden an VCLP bzw. VURS übermittelt. Die Nutzerinnen und Nutzer können selbst einstellen, wieviele Informationen in ihrem Profil für andere sichtbar sein sollen.

Dezentrale Architektur der Lernumgebungen

Bei der Konzeption der VCLP war am KIT von Anfang an daran gedacht worden, keinen ausschließlich zentralen Ansatz zu verfolgen. Eine zentrale EPICUR-Lernplattform ohne weitere Integration hätte zur Folge gehabt, dass die Kursanbieter gezwungen wären, ihre auf den lokalen Lernplattformen bestehenden Kurse neu zu erstellen. Außerdem wäre eine Einarbeitung in die neue Umgebung für die Kursanbieter notwendig gewesen, insbesondere wenn im Umgang mit ILIAS noch keine Erfahrung besteht. In Konsequenz hätten die meisten Kursinhalte dann parallel auf zwei Lernplattformen – einmal auf der lokalen Plattform und einmal auf der EPICUR-Plattform – aktuell gehalten werden müssen.

Um diese Probleme zu vermeiden, wurde ein kombinierter Plattform-Ansatz verfolgt, der sowohl eine zentrale Lernplattform als Virtual Campus Learning Platform bereitstellt, als auch die bereits bestehenden Lernplattformen an den Partneruniversitäten für EPICUR-Kurse integriert. Kursanbieter haben somit die Wahl, ob sie ihren Kurs auf der zentralen virtuellen Plattform anbieten möchten oder weiterhin auf ihrer gewohnten lokalen Lernplattform. Die virtuelle Plattform ist aber immer das „Gateway“ für die EPICUR-Kursteilnehmenden, egal auf welcher Plattform der Kurs tatsächlich läuft.

Middleware

Das SCC hat hierzu einen E-Learning Community Server (ECS) aufgesetzt, der als Middleware agiert und Informationen mittels eines nachrichtenorientierten Ansatzes zwischen den angeschlossenen Lernplattformen austauscht. Die Sicherheit und Integrität beim Datenaustausch zwischen den Lernplattformen und dem ECS wird durch vom SCC ausgestellte Client- und Serverzertifikate sichergestellt. Die Integration der ILIAS-Lernplattformen der Universität Freiburg und des KIT waren einfach möglich, da ILIAS bereits

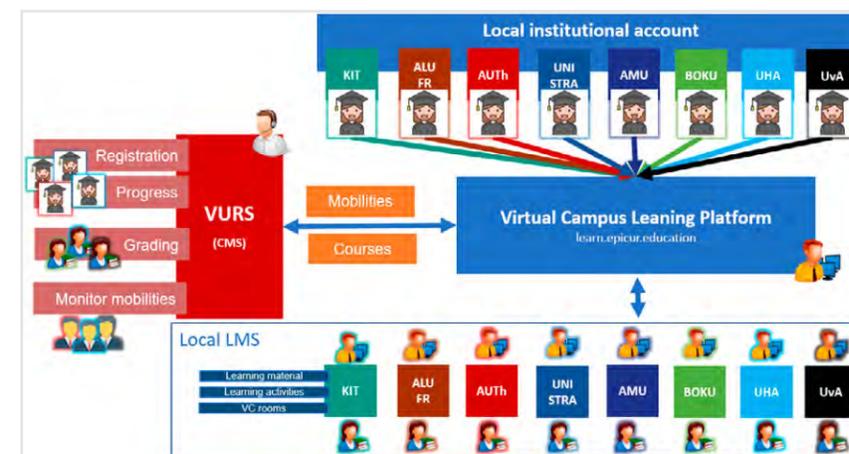


Abbildung 2: Der EPICUR Inter-University Campus besteht aus zwei Systemen.

eine native Unterstützung für eine ECS-Anbindung mitbringt. Für die anderen Partneruniversitäten, die Moodle als Lernplattform einsetzen, wurde ein entsprechendes Moodle-ECS-Plugin installiert und für die ECS-Kommunikation eingerichtet. Die Universität von Amsterdam (UvA) setzt als Lernplattform Canvas ein, hier entwickeln SCC und UvA gemeinsam ein Konzept, um eine Konnektivität von Canvas mit dem ECS mittels eines ECS-Proxy zu realisieren.

Komfort beim Kursmanagement

Möchte ein Kursanbieter seinen Kurs auf der VCLP zur Verfügung stellen, muss er nur in den Kurseinstellungen die Freigabe für EPICUR aktivieren. Wenige Augenblicke später erscheint dann der Kurs auf der VCLP-Plattform und kann von angemeldeten und für diesen Kurs freigeschalteten Kursteilnehmerinnen und -teilnehmern aufgerufen werden. Beim Klick auf den Kursnamen aus VCLP heraus wird man auf die lokale Lernplattform weitergeleitet und direkt angemeldet. Es werden dazu lokale Accounts auf den Lernplattformen in der Form EPICUR_<EPPN> erstellt. Diese Accounts sind allerdings nur für die Lernplattform nutzbar und keine vollwertigen Uni-Accounts, für die eine aufwändige Immatrikulation notwendig wäre (Abbildung 3).

Durch die Integration der virtuellen Lernplattform VCLP mit dem zentralen Campus-Management-System VURS wird sichergestellt, dass nur die Kurse zu sehen sind, für die man sich angemeldet hat und dafür auch zugelassen wurde. Über das VURS können Kursanbieter Kursteilnehmenden auch das Passwort für den Beitritt zum Kurs mitzuteilen, wenn sie dieses wünschen.

Ausblick

Für eine zweite EPICUR-Phase haben sich die Partner vorgenommen, auch die Campus-Management-Systeme stärker miteinander zu vernetzen. Ein wichtiger Schritt dazu sind die Initiativen auf europäischer Ebene, die die gemein-

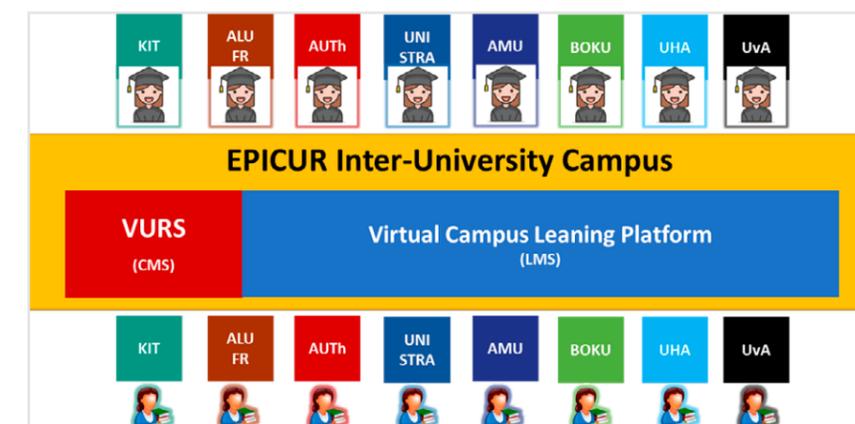


Abbildung 3: VURS, VCLP und lokale Lernmanagementsysteme sind integriert und mittels des Accounts der Heimatuniversität zugänglich.

samen Datenaustauschformate ELMO und EDCI entwickeln und spezifizieren. Entsprechende Unterstützung in den bestehenden Campus-Management-Systemen vorausgesetzt, können die an einer EPICUR-Universität erworbenen Leistungen eines Studierenden digital mit der Heimat-Universität ausgetauscht werden und die Anerkennungsprozesse IT-gestützt ablaufen.

EPICUR Inter-University Campus

With the Virtual Campus Learning Platform (VCLP), SCC has solved an important task in teaching in the EPICUR Alliance. The VCLP provides a central gateway to the entire EPICUR course offering and integrates the existing learning platforms at the eight partner universities. Students and teachers can easily log in using their existing university accounts and directly participate in the EPICUR courses offered on the different teaching platforms.

Transkription für frühneuzeitliche Drucke – vom Bilddigitalisat zum digitalen Volltext

Erfolgreicher Abschluss von Phase II des von der DFG geförderten Projekts ‚OCR-D‘, dessen vorrangiges Ziel die konzeptionelle und technische Vorbereitung der Volltextdigitalisierung für frühneuzeitliche Drucke ist.

Volker Hartmann

Motivation

OCR-D¹ ist ein von der Deutschen Forschungsgemeinschaft (DFG) gefördertes Projekt zur Weiterentwicklung von Verfahren der Optical Character Recognition (OCR) für frühneuzeitliche Drucke. Die Transkription von Bilddigitalisaten, die Umwandlung in (digitalen) Volltext, ist für Drucke ab dem 20. Jahrhundert praktisch gelöst. Viele Bibliotheken stellen Bilddigitalisate und deren Volltexte (teilweise gegen Bezahlung) in hoher Qualität zur Verfügung. In den vergangenen 15 Jahren wurde, gefördert von der DFG, eine Bilddigitalisierung für einen beträchtlichen Teil des in Deutschland gedruckten Materials aus der Zeit von 1500 bis ca. 1850 durchgeführt. Um die alten Drucke zu bewahren, ist der physikalische Zugang zu den Werken oft eingeschränkt. Durch die Digitalisierung sind Forschende nun in der Lage, alte Bücher aus verschiedenen Bibliotheken direkt am PC zu vergleichen. Der nächste logische Schritt ist die inhaltliche Analyse der Bücher (z.B. um damit die Frage zu beantworten, welche Schreibweisen wo und wann das erste Mal verwendet wurden und wie sie sich ausgebreitet haben). Jedoch stellt die Volltextdigitalisierung dieser frühneuzeitlichen Drucke eine große Herausforderung dar. Bis Mitte des 19. Jahrhunderts existierte kein Standard für die Schriftarten, und die Schreibweise von Wörtern war noch nicht normiert, um nur zwei der Schwierigkeiten aufzuführen. Das Projekt OCR-D hat zur Aufgabe, die Lücke vom digitalen Bild bis zum digitalen Volltext zu schließen und ein Konzept zu erarbeiten, mit dessen Hilfe die Bibliotheken die vorliegenden digitalen Bilder transkribieren können.

Das Projekt ‚OCR-D‘

In der ersten Phase des seit 2015 von der DFG geförderten Projekts stand die Bestimmung von Verarbeitungsschritten innerhalb der Verarbeitungskette im Mittelpunkt. Dazu wurde ein feingranulares Funktionsmodell entworfen. Der gesamte Prozess wurde in mehrere, nacheinander auszuführende Module (Abbildung 1) aufgeteilt. Ein Modul (z.B. Bildvorverarbeitung) besteht aus einem oder mehreren Prozessoren für die Teilaufgaben Zuschneiden, Binarisierung, ..., die bei Bedarf eine Teilaufgabe des Moduls übernehmen. Für alle Prozessoren wurde eine einheitliche Schnittstelle entwickelt und die jeweiligen Ein- und Ausgangszustände definiert. Im Laufe dieser ersten Phase ergänzte das SCC die Kompetenzen von OCR-D im Bereich Softwaretechnologie mit dem Ziel, die technische Realisierbarkeit und die Interoperabilität zwischen den Prozessoren sicherzustellen. Für die Implementierung der Module wurden jeweils Modulprojekte von der DFG ausgeschrieben, die dann in Phase II des Projekts von der Berlin-Brandenburgischen Akademie der Wissenschaften, der Herzog-August Bibliothek Wolfenbüttel, der Staatsbibliothek zu Berlin sowie dem KIT bzw. SCC koordiniert wurden. Durch die Aufteilung in unabhängige Module kann jedes Modul speziell auf die Anforderungen für frühneuzeitliche Drucke angepasst werden. Da auch zum Teil konkurrierende Implementierungen eines Moduls gefördert wurden, ist es möglich, diese entsprechend ihrer Stärken einzusetzen. Das erste Modul schneidet den Rand des Bilddigitalisats ab und wandelt es in ein binäres Bild um. Das zweite Modul detektiert und kennzeich-

net die Textbereiche. Im dritten Modul wird die Zeichenerkennung durchgeführt, dessen Ergebnis dann mithilfe des vierten Moduls bei Bedarf korrigiert wird. Das letzte Modul speichert die (Zwischen-)Ergebnisse im Repository ab. Speziell für die Zeichenerkennung (kurz OCR) wurden für die häufig verwendeten Schriftarten die OCR-Modelle neu eingelernt. Die Nachkorrektur versucht, anhand der erkannten Buchstaben/Wörter und des Kontexts, z.B. mittels für den jeweiligen Erscheinungszeitraum der Drucke erstellten Lexika, falsch erkannte Zeichen zu korrigieren. Mit dem letzten Modul werden die Daten zusammen mit den (Zwischen-)Ergebnissen und der Provenienz (Logeinträge für die Reproduzierbarkeit der Ergebnisse) in einen Container gepackt und im Forschungsdatenrepository gespeichert. Wie für alle Entwicklungen rund um OCR-D ist die Verfügbarkeit der Implementierungen in GitHub² unter einer offenen Lizenz ein wesentlicher Aspekt. Damit die Module von den Bibliotheken verwendet werden können, ist außerdem wichtig, dass die darin verwendeten Algorithmen keine besonderen Anforderungen an die Hardware stellen. So muss ein Modul auf einem handelsüblichen PC ausführbar sein und, im Vergleich zu bekannten (kommerziellen) Tools, nicht übermäßig viel Rechenzeit beanspruchen.

Das im Rahmen des Projekts entwickelte OCR-D-Framework stellt eine Umgebung bereit, mit der ein kompletter Transkriptionsworkflow ausgeführt werden kann. Dieses beinhaltet alle verfügbaren und getesteten Module, eine Workflow Engine (Taverna) zum Ausführen

¹ ocr-d.de
² github.com/OCR-D

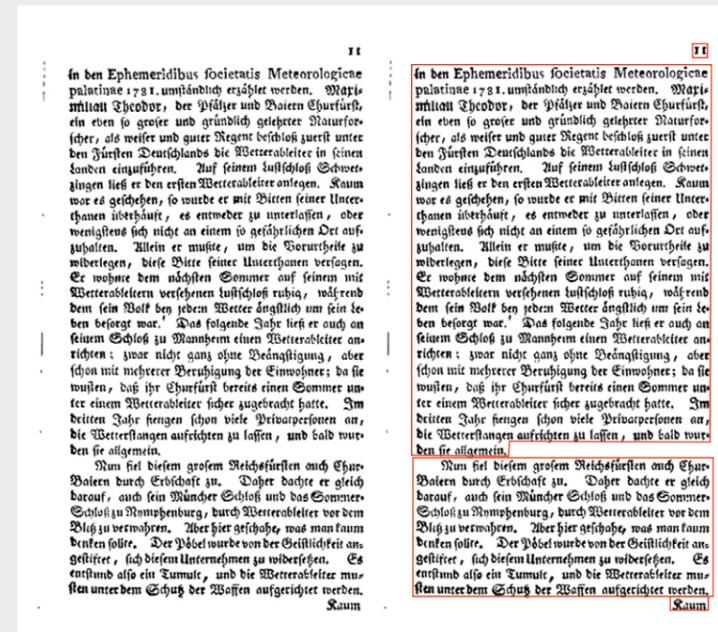


Abbildung 1: Einfacher Transkriptionsworkflow vom digitalen Bild zum digitalen Volltext: Bilddigitalisat (oben) → Bildvorverarbeitung: Binarisiertes Bild (unten links) → Texterkennung: Erkannte Textbereiche (unten Mitte) → OCR: Erkannte Buchstaben (fehlerhafter erkannte Buchstaben sind rot eingefärbt) (unten rechts) → Nachkorrektur → Forschungsdatenrepositorium

Quelle: Gescannte Seite von ‚Unterricht vom Blitz und den Blitz- oder Wetter-Ableitern‘, Johann Friedrich Luz, 1784, [purl: deustextarchiv/delluz_bltz_1784](http://purl.de/deustextarchiv/delluz_bltz_1784)

des Transkriptionsworkflows und ein Forschungsdatenrepositorium, in dem die (Zwischen-)Ergebnisse gespeichert werden können. Die Komponenten des OCR-D-Frameworks können einzeln installiert und konfiguriert oder komplett als Komposition mehrerer Docker-Container installiert werden. Am Ende des Transkriptionsworkflows werden alle (Zwischen-)Ergebnisse und die Provenienz (PROV-XML³) im Forschungsdatenrepositorium

gespeichert. Anschließend kann über den Browser auf die Ergebnisse und deren Metadaten zugegriffen werden.

Neben dem Forschungsdatenrepositorium, welches standardmäßig nur lokal zugreifbar ist, wurde noch ein öffentliches Ground-Truth-Repository⁴ aufgebaut. Von diesem werden ausgewählte Beispiele von frühneuzeitlichen Drucken mit den erwarteten und manuell kontrollierten



in den Ephemeridibus societatis Meteorologicae palatinae 1781, umständlich erzählt werden. Maximilian Theodor, der Pfälzer und Baiern Churfürst, ein eben so großer und gründlich gelehrter Naturforscher, als weiser und guter Regent beschloß zuerst unter den Fürsten Deutschlands die Wetterableiter in seinen Landen einzuführen. Auf seinem Luftschloß Schwetzingen ließ er den ersten Wetterableiter anlegen. Kaum war es gechehen, so wurde er mit Bitten seiner Unterthanen überhäuft, es entweder zu unterlassen, oder wenigstens sich nicht an einem so gefährlichen Ort aufzuhalten. Allein er mußte, um die Vorurtheile zu widerlegen, diese Bitte seiner Unterthanen verjagen. Er wohnte dem nächsten Sommer auf seinem mit Wetterableitern versehenen Luftschloß rubia, während dem sein Volk den jedes Wetter ängstlich um sein Leben besorgt war. Das folgende Jahr ließ er auch an seinem Schloß zu Mannheim einen Wetterableiter anrichten: zwar nicht ganz ohne Bedäunigung, aber schon mit mehrerer Beruhigung der Einwohner; da sie wollten, daß ihr Churfürst bereits einen Sommer unter einem Wetterableiter sicher zugebracht hatte. Im dritten Jahr fingen schon viele Privatpersonen an, die Wetterlängen aufzichten zu lassen, und bald wurden sie allgemein.

Nun ist diesem großen Reichthum auch Churfürstlichen durch Erbschaft zu. Daher dachte er gleich darauf, auch sein Müncher Schloß und das Sommer-Schloß zu Mannheim, durch Wetterableiter vor dem Blitz zu verwahren. Aber hier geschah, was man kaum denken sollte. Der Pöbel wurde von der Geilichkeit angegriffen, sich diesem Unternehmen zu widersetzen. Es entstand also ein Tumult, und die Wetterableiter mußten unter dem Schuß der Waffen aufgerichtet werden. Kaum

in den Ephemeridibus societatis Meteorologicae

[...]
den Fürsten Deutchlands die Wetterableiter in feinen Landen einzuführen. Auf seinem Luftschloß Schwetzingen ließ er den ersten Wetterableiter anlegen. Kaum [...]

³ www.w3.org/TR/prov-xml/
⁴ ocr-d-repo.scc.kit.edu/

Mit Hilfe des Frameworks und der Ground-Truth-Daten konnten die Module von uns evaluiert und gleichzeitig die Funktionsfähigkeit des Frameworks auf ‚normalen‘ Rechnern gezeigt werden. Die Rechenzeit lag, abhängig von der Komplexität der zu transkribierenden Seite und den verwendeten Modulen, zwischen 10s und 30s. Für einfache Seiten konnte mit den für OCR-D gemachten Anpassungen eine Fehlerrate von unter 1%, was

ungefähr einem falsch erkannten Zeichen in zwei Zeilen entspricht, erreicht werden. Für komplexere Seiten (z.B. mit verschiedenen Schriftgrößen/Schriftarten) konnte immer noch eine Fehlerrate von unter 10% erzielt werden.

In der bereits gestarteten Phase III des Projekts kann nun, aufbauend auf dem vorhandenen OCR-D-Framework, die Integration in den Bibliotheken vorange-

trieben werden. Durch den breiten Einsatz mit verschiedensten Drucken sollen mögliche Schwachpunkte besser erkannt und die Prozessoren und der Workflow weiter optimiert werden. Schulungen und Vorträge werden die weitere Verbreitung unterstützen.

Transcription for Early Modern Prints – From Image Digitization to Digital Full Text

OCR-D is a DFG-funded initiative for Optical Character Recognition (OCR) development for early modern prints. While OCR for prints from the 20th century onwards has been solved, early modern prints still remain a great challenge. The main goal of OCR-D is the conceptual and technical preparation of full text digitization for these early modern prints. For this purpose, the OCR was divided into several modules, each of which was then optimized for the specific requirements. Using the OCR-D framework developed as part of the project, the modules could be evaluated with manually produced results from early modern prints (Ground Truth Repository). An error rate of less than 1% was achieved for pages with a simple layout. In the next phase of the project, the integration in the libraries will be pushed forward and the modules will be further optimized.

Diese virtuelle Veranstaltungsreihe richtete sich an alle RSEs und auch jene, die sich im Herzen bereits als RSE definieren, aber noch nicht in den Genuss dieser offiziellen Stellenbeschreibung kommen. Die Veranstaltungen sollten RSEs die Gelegenheit bieten, trotz und gerade auch wegen COVID-19 ihre Expertise zu entwickeln und weiter auszubauen, sich mit Expertinnen und Experten zu vernetzen, neue Kollaborationen einzugehen und sich weltweit mit Gleichgesinnten auszutauschen. All dies natürlich in einer möglichst offenen, inspirierenden und angenehmen Atmosphäre.

Das Konzept von SORSE baute auf inhaltlicher sowie zeitlicher Flexibilität auf und ermöglichte monatlich fortlaufende Einreichungen zu den unterschiedlichsten Formaten: von Talks, über Workshops, Software-Demonstrationen, Panels und Diskussionen sowie Postern bis hin zu

Blog Posts. Die zahlreichen großartigen Beiträge von RSEs weltweit haben es ermöglicht, mehr als sieben Monate lang eine wöchentliche Veranstaltung anzubieten. Das abwechslungsreiche Themenspektrum richtete sich an RSEs aus unterschiedlichen Bereichen und Karriere-stufen: von Research Software Directories, FAIRness von Forschungssoftware, dem Erreichen der Produktreife mit Forschungssoftware, der gezielte Umgang mit wechselndem Personal in Forschungsprojekten bis hin zum Aufbau von Research Software Communities; aber auch Themen aus spezifischen Domänen wie zum Beispiel Archäologie, den Geisteswissenschaften oder der Bioinformatik. Dieses sehr abwechslungsreiche Programm fand nicht nur großen Zuspruch, sondern hat regelmäßig neue wie wiederkehrende Teilnehmerinnen und Teilnehmer begeistert.

SORSE war rückblickend ein echter Gewinn, nicht nur für das Organisationsteam, sondern vor allem auch für die in der bzw. für die Forschung entwickelnden Informatikerinnen und Informatiker, Softwareentwicklerinnen und -entwickler, RSEs und Wissenschaftlerinnen bzw. Wissenschaftler. Alle Teilnehmenden konnten nicht nur internationale Bande knüpfen und stärken, sondern haben vor allem auch als Gemeinschaft profitiert und enorm viel gewonnen. Die Mitarbeit bei der Organisation fühlte sich nicht als zusätzliche Arbeit an; sie wurde von allen Beteiligten viel eher als bereichernd wahrgenommen, insbesondere, weil es die Möglichkeit gab das Bewusstsein für die Relevanz von RSEs auch im SCC zu schärfen. Die Verantwortlichen sind schon sehr gespannt, was die Zusammenarbeit der unterschiedlichen RSE-Initiativen in Zukunft noch bringen wird und wie dies weiterhin tatkräftig unterstützt werden kann.

Rückblick auf großartige Monate mit SORSE

Seit der COVID-19-Pandemie läuft zugegebenermaßen so einiges anders. Trotzdem gibt es gerade durch die Digitalisierung und den mittlerweile sehr selbstverständlichen Einsatz von Videokonferenztools umfangreiche Möglichkeiten, um großartige digitale und auch interaktive Veranstaltungen zu organisieren und durchzuführen. Mit der Series of Online Research Software Events, kurz SORSE, wurde genau das gemacht und über sieben Monate wöchentlich ein abwechslungsreiches Format zu den unterschiedlichsten Themenfeldern für und von Research Software Engineers geboten.

Eileen Kühn, Stephan Druskat¹

Es ist, als wäre es gestern gewesen, dass die Organisation der deRSE Konferenz 2020 – eine Konferenz zum Thema Forschungssoftware in Deutschland – begann. Die Planung übernahm ein Organisationskomitee von mehr als 20 Freiwilligen aus verschiedenen Universitäten und Forschungseinrichtungen in Deutschland, zusammengerufen durch die de-RSE e.V. – Gesellschaft für Forschungssoftware². Das Komitee war guter Dinge und hatte

schon einige Aufgaben auf den To-do-Listen abgearbeitet. Doch gleichzeitig war es auch die Zeit der großen Ungewissheit, geprägt durch die COVID-19-Pandemie. Daher wurde entschieden, keine Präsenzveranstaltung, sondern stattdessen eine virtuelle Veranstaltung durchzuführen. Dies eröffnete die Gelegenheit, auch andere nationale Initiativen von Research Software Engineers (RSEs) – aus dem Vereinigten Königreich, den Niederlan-

den, Belgien, Skandinavien, den USA, Australien und Neuseeland – an Bord zu holen, um gemeinsam die internationale Veranstaltung zu organisieren.

So lautete schließlich die internationale Antwort auf COVID-19: A Series of Online Research Software Events (SORSE).

¹ Deutsches Zentrum für Luft- und Raumfahrt (DLR)

² de-rse.org

Wer ist SORSE?

Die Veranstaltungsreihe SORSE wurde ausschließlich von Freiwilligen aus den unterschiedlichen internationalen RSE-Initiativen, unter anderem aus Deutschland, dem Vereinigten Königreich, den USA, den Niederlanden, den nordischen Ländern, Australien und Neuseeland sowie Belgien, auf die Beine gestellt und organisiert. Eine vollständige Liste der Verantwortlichen und ihrer jeweiligen Aufgaben und Zugehörigkeiten kann auf der SORSE-Webseite eingesehen werden. Ein großer Dank gebührt natürlich auch den zahlreichen Vortragenden sowie den Teilnehmerinnen und Teilnehmern, ohne die diese Veranstaltungsreihe nie so spannend, lebendig und erfolgreich geworden wäre. Auf (fast) alle Beiträge kann auch jetzt noch über die Veranstaltungswebseite⁴ zugegriffen werden.

Checkliste für Research Software Engineers⁵

Ein Research Software Engineer (RSE) kombiniert professionelles Software-Engineering-Know-how mit einem tiefen Verständnis für die Forschung. Im Rahmen einer engen Zusammenarbeit mit Forschenden entwickeln, pflegen und erweitern RSEs Software, um computergestützte wissenschaftliche Forschung zu ermöglichen und voranzutreiben.

Wenn auf Sie einige der folgenden Fragen zutreffen, arbeiten vermutlich auch Sie als Research Software Engineer:

- Entwickeln Sie Software um zu forschen?
- Verbringen Sie mehr Zeit mit der Entwicklung von Software als mit der Forschung?
- Werden Sie manchmal nicht als Autor in Papern benannt, obwohl Sie eine wesentliche Rolle bei der Entwicklung der Software spielen?
- Fehlen Ihnen die für Ihre akademische Karriere notwendigen Kennzahlen, obwohl Sie einen wesentlichen Beitrag durch Software geleistet haben?

Looking back on great months with SORSE

Things have been different since COVID-19. Nevertheless, digitalization and the use of video conferencing tools, which is now taken for granted, have given us extensive opportunities to organize and hold great digital and interactive events. With the Series of Online Research Software Events (SORSE) we did exactly that and offered a diverse format on different topics for and by research software engineers every week for more than 7 months. If you missed the chance to attend you can still access the recordings of almost all contributions via the SORSE event website⁵.

³ sorse.github.io/contact/committees/

⁴ sorse.github.io/programme/

⁵ sorse.github.io/programme/

Das EuroHPC-Projekt MICROCARD stellt sich vor

Im von der Europäischen Union geförderten EuroHPC-Projekt MICROCARD entwickelt ein Team aus elf Institutionen eine Software, mit der kardiovaskuläre Simulationen auf zukünftigen Exascale-Supercomputern ausgeführt werden können. Die dafür benötigten Algorithmen müssen auf das jeweilige mathematische Problem, den Umfang der Berechnungen und das besondere Design dieser zukünftigen Computer zugeschnitten sein.

Hartwig Anzt

Herz-Kreislauf-Erkrankungen sind weltweit die häufigste Todesursache und die Hälfte dieser Todesfälle ist auf Herzrhythmusstörungen, also Störungen des elektrischen Synchronisationssystems des Herzens, zurückzuführen. Computermodelle sind unerlässlich, um das Verhalten dieses komplexen Systems und seiner Krankheiten zu verstehen. Diese Modelle sind bereits sehr ausgereift und weit verbreitet, aber derzeit sind sie noch nicht leistungsfähig genug, um die (2 Milliarden!) Einzelzellen des Herzens zu berücksichtigen. Sie müssen daher davon ausgehen, dass Hunderte von Zellen sich ungefähr gleich verhalten. Aufgrund dieser Einschränkung können aktuelle Modelle die Ereignisse in alternierenden und strukturell erkrankten Herzen nicht darstellen, bei denen eine reduzierte elektrische Kopplung zu großen Verhaltensunterschieden zwischen benachbarten Zellen mit möglicherweise fatalen Folgen führt.

Wenn wir das Herz Zelle für Zelle modellieren wollen, stehen wir vor einem 10.000-fach größeren und auch schwerer zu lösenden mathematischen Problem. Wir brauchen leistungsfähigere Supercomputer als die heute existierenden und viel Erfindungsreichtum, um unser Problem auf diesen zukünftigen Maschinen effizient zu lösen.



Angesichts der Bedeutung der kardiovaskulären Forschung fördert die Europäische Union die Entwicklung einer Software, mit der kardiovaskuläre Simulationen (Abbildung 1) auf zukünftigen Exascale-Supercomputern ausgeführt werden können. Im Projekt MICROCARD wird ein Team von elf Institutionen (Abbildung 2) Algorithmen

entwickeln, die auf das jeweilige mathematische Problem, den Umfang der Berechnungen und das besondere Design dieser zukünftigen Computer zugeschnitten sind, die wahrscheinlich den größten Teil ihrer Rechenleistung mit ultra-parallel Computer Elementen wie Grafikprozessoren erbringen. Das Team strebt keinen „Proof of Concept“ an, sondern wird den entwickelten Code direkt produktionsreif machen, um damit reale Probleme in der Kardiologie zu lösen. Daher umfasst das Projekt Expertinnen und Experten der Computerwissenschaft, der Mathematik und der Biomedizintechnik und arbeitet mit Fachleuten aus Kardiologie und Physiologie zusammen. Die Forschungsgruppe FINE¹ am SCC spielt eine zentrale Rolle im EuroHPC-Projekt MICROCARD: Sie wird den numerischen Kern des Exascale-fähigen Software-Stacks entwickeln und einsetzen sowie die Plattformportabilität über die Open-Source-Bibliothek Ginkgo ermöglichen.

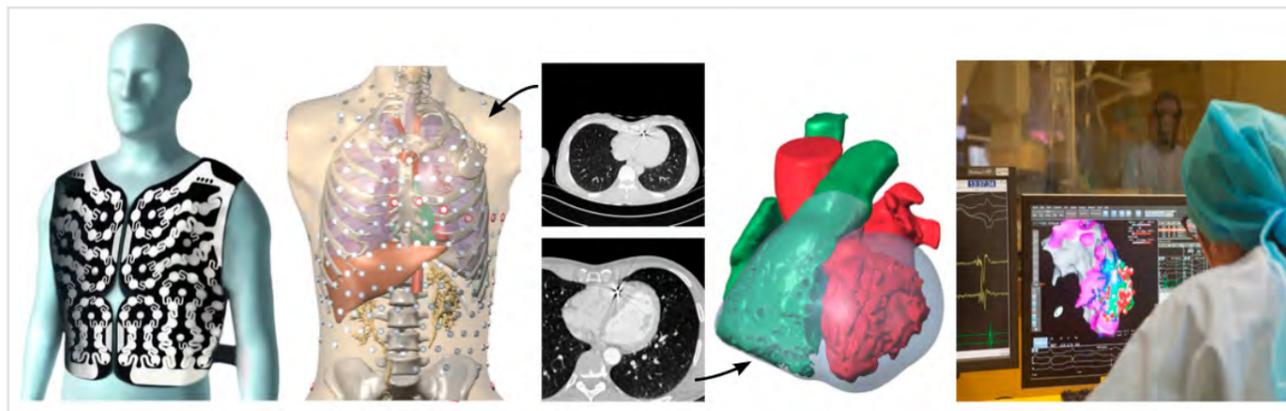


Abbildung 1: Mit Hilfe von Elektroden werden bei einem EKG die elektrischen Ströme gemessen, die das menschliche Herz schlagen lassen (links); eine Diskretisierung erlaubt es, einen digitalen Zwilling des menschlichen Herzens zu erstellen (Mitte); zur Simulation der kardioelektrischen Effekte werden hochintelligente Algorithmen und leistungsfähige Supercomputer benötigt (rechts). (Copyright by Gunnar Seemann)

¹ www.scc.kit.edu/ueberuns/jrg-fine.php

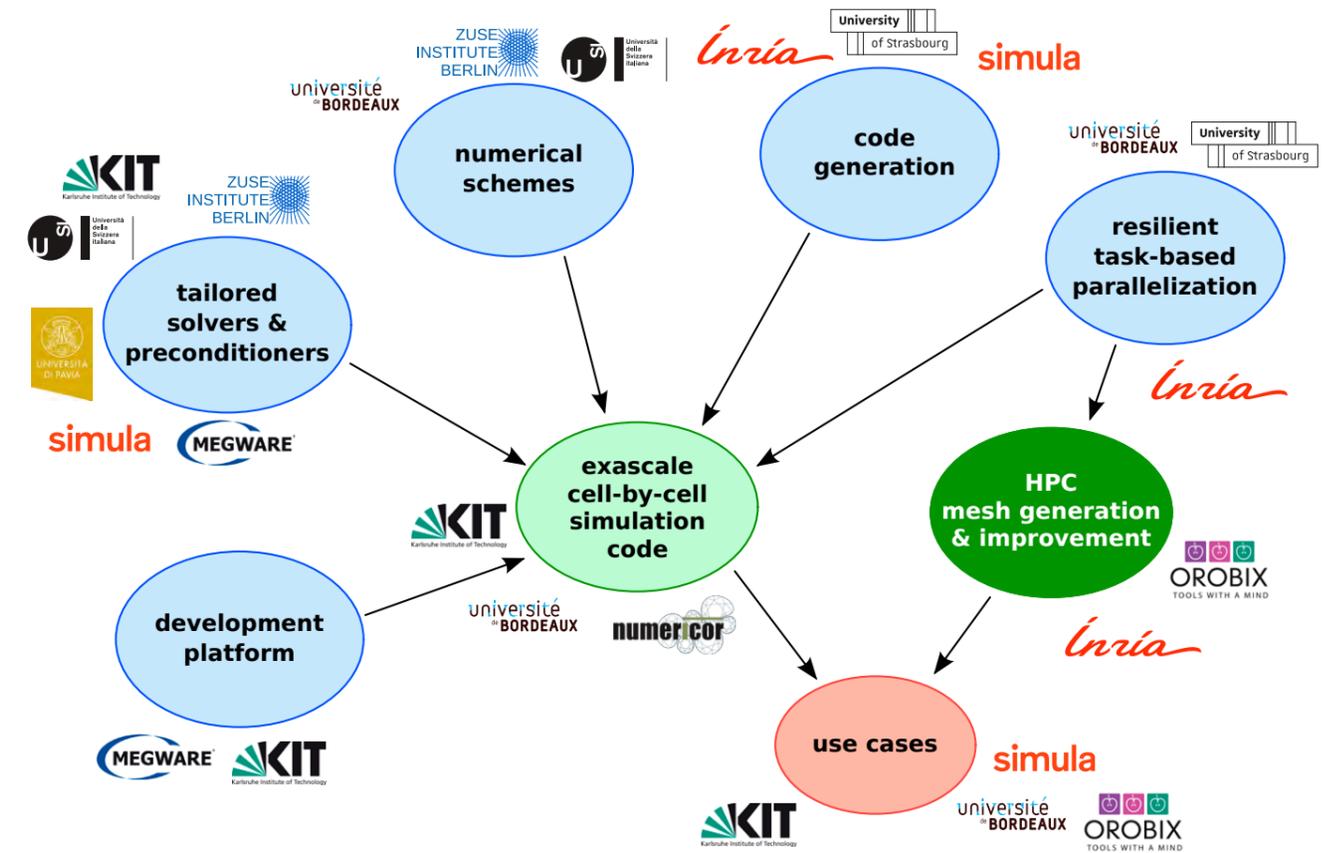


Abbildung 2: Die beteiligten Institutionen und deren unterschiedliche Beiträge und Schwerpunkte zum Projekt MICROCARD.

Introducing the EuroHPC project MICROCARD

Cardiovascular diseases are the most frequent cause of death worldwide and half of these deaths are due to cardiac arrhythmia, disorders of the heart's electrical synchronization system. Computer models are essential to understand the behaviour of this complex system and its diseases. These models are already very sophisticated and widely used, but currently they are not powerful enough to take the heart's (2 billion!) individual cells into account. They must therefore assume that hundreds of cells are doing approximately the same thing. Due to this limitation, current models cannot represent the events in aging and structurally diseased hearts, in which reduced electrical coupling leads to large differences in behaviour between neighbouring cells, with possibly fatal consequences.

If we want to model the heart cell by cell, we face a mathematical problem that is 10,000 times larger, and also harder to solve. We will need larger supercomputers than those that exist today, and a lot of inventiveness to solve our problem efficiently on these future machines.

Given the relevance of cardiovascular research, the European Union has decided to develop software capable of running cardiovascular simulations on future exascale supercomputers. In the MICROCARD project, a team of 11 institutions will develop algorithms that are tailored to the specific mathematical problem, to the size of the computations, and to the particular design of these future computers, which will probably owe most of their compute power to ultra-parallel computing elements such as Graphics Processing Units. The team does not aim at a "proof of concept", but will turn the developed code production-ready to solve real-life problems in cardiology. Therefore the project includes computer experts, mathematicians, biomedical engineers, as well as cardiologists and physiologists.

The research group FINE¹ at SCC plays a central role in the EuroHPC project: They will develop and deploy the numerical core of the exascale-ready software stack and enable platform portability via the Ginkgo open source library.

Maschinelles Lernen / Künstliche Intelligenz in Biosimulationen

Universität Duisburg/Essen, Forschungszentrum Jülich, Deutsches Krebsforschungszentrum und das SCC entwickeln eine selbstanpassende Variante der dynamischen Partikelschwarmoptimierung von Biomolekülsimulationen – Paper in Nature Machine Intelligence veröffentlicht.

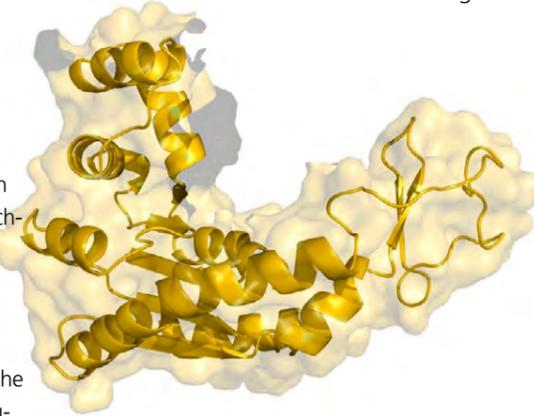
Alexander Schug

Leben funktioniert auf der molekularen Skala durch ein komplexes Zusammenspiel von Biomolekülen, für welches die Struktur der beteiligten Biomoleküle eine zentrale Rolle spielt. Experimentelle Methoden können solche Strukturen bestimmen und beruhen auf der Auswertung der Primärdaten, deren Interpretation aber nicht immer eindeutig ist. Molekulare Simulationen sind ein mächtiges Werkzeug, um solche mehrdeutigen experimentellen Daten auszuwerten.

Eine essentielle Herausforderung besteht darin, die experimentelle Information gegen das zugrunde liegende physikalische Simulationsmodell zu gewichten. In Zusammenarbeit von FZJ, KIT, DKFZ und der Universität Duisburg/Essen wurde eine

selbstanpassende Variante der dynamischen Partikelschwarmoptimierung entwickelt, um dieses Gewichtungproblem zu lösen. Jeder Einzelparameter wird zur Laufzeit gelernt, was eine sich dynamisch entwickelnde und iterativ verfeinerte Suchraumtopologie ergibt. Dabei konnte für mehrere biomolekulare Systeme gezeigt werden, dass die Methode die Rechenzeit sehr effizient nutzt und gleichzeitig hoch-akkurate Strukturen liefert. Da solche Parameterprobleme in molekularen Simulationen häufig sind, sind neben biomolekularen Simulationen auch Anwendungen wie z.B. Materialsimulationen denkbar.

Paper: doi.org/10.1038/s42256-021-00366-3
 News des KIT: www.kit.edu/kit/29451.php



Biomoleküle sind die molekularen Schweizer Taschenmesser in unserem Körper und führen unterschiedlichste Funktionen aus. Das abgebildete Protein reguliert als Enzym den Energieträger ATP. (Grafik: Alexander Schug, KIT)

Intelligentes Batteriemangement durch Spektroskopie und Maschinelles Lernen

Forschungszentrum Jülich, Fritz-Haber-Institut Berlin und das SCC entwickeln im Projekt i2Batman ein neuartiges intelligentes Batteriemangementssystem, das die Ladezyklen, basierend auf einer KI und einem digitalen Zwilling der Batterie, optimiert.

Ivan Kondov

Das Schnellladen und die Lebensdauer von elektrischen Batterien sind wichtige Voraussetzungen für die breitere Anwendung von Elektrofahrzeugen. Hierzu bietet das sogenannte Batteriemangementssystem ein großes Optimierungspotential. Da über die Zusammenhänge zwischen diesen zwei Anforderungen und den internen Batterieparametern noch zu wenig bekannt ist, gehen die aktuell eingesetzten Batteriemangementssysteme „auf Nummer sicher“ und setzen häufig unnötig strenge Sicherheitsbeschränkungen für den Betrieb der Batterien.

Zusammen mit Partnern am Forschungszentrum Jülich und am Fritz-Haber-Institut Berlin hat sich das SCC als Ziel gesetzt, ein neuartiges intelligentes Batteriemangementssystem zu entwickeln, das anhand

eines detaillierten Ersatzmodells (digitaler Zwilling) der Batterie und einer Künstlichen Intelligenz (KI) bessere Entscheidungen zu den Ladezyklen der Batterie treffen kann. Die Aufgabe im Projekt ist es, dieses Batterieersatzmodell auf der Ebene der einzelnen Batteriezellen mithilfe des Gaußprozessverfahrens zu entwickeln. Dabei wird das Modell mit Parametern verschiedener Ersatzschaltungen und dem Ladezustand jeder Zelle parametrisiert. Das Modell wird sowohl mit experimentellen Spektroskopiedaten als auch mit Daten von physikalischen Ersatzschaltungsmodellen trainiert. Zudem ermöglicht der Einsatz des Gaußprozessverfahrens eine Bestimmung

der Modellunsicherheit (Uncertainty Quantification), welche für die Funktion der KI erforderlich ist.

In einer ersten Testphase wird das mit der KI ausgestattete Batteriemangementssystem auf einfacher Hardware implementiert. Dabei werden im laufenden Betrieb noch Daten gesammelt und gespeichert, so dass sich die KI weiter verbessern kann. In der finalen Phase wird das System mit einer vollständig charakterisierten Batterie getestet. Das Projekt i2Batman ist eines von 19 geförderten Projekten der Helmholtz Artificial Intelligence Cooperation Unit (helmholtz.ai).



Laden eines E-Autos (Possessed Photography on Unsplash)

wfGenes: Boosting productivity of workflow developers

Rapid design of workflow and task-based parallel applications is a major challenge in computational materials science. Limiting issues are the low productivity of development due to a limited number of reusable components and the lacking portability of the applications between different workflow management systems and parallel computing platforms. Within a project of RTG 2450 Computational Nanoscience, the lightweight tool wfGenes has been provided to tackle these issues based on an automated generative approach.

Mehdi Roozmeh and Ivan Kondov

Computational materials research is important for design of novel materials with predefined properties. Compared to other application domains of computational science it has specific challenges mainly caused by a wide range of software modules that have to be integrated to perform a study. A particular software module fulfills a specific task in the whole simulation that is often represented in the form of a scientific workflow. To steer such complex simulations on supercomputers, specialized workflow management systems are used. Due to diverse application requirements, a great variety of such workflow systems currently exist [1]. In the recent years, we have experimented with several different workflow management systems, e.g. with UNICORE [2] and, more recently, we have successfully employed FireWorks [3] to perform virtual screening of cathode catalysts for fuel cells [4] and anode catalysts for water hydrolyzers [5].

Nevertheless, the current best practices still require advanced IT skills and substantial effort for the development,

verification and running of workflow applications. In addition, the portability of the artifacts to other applications and user groups is limited due to the concrete workflow management system or workflow language used. Another lesson learned is that one workflow management system cannot satisfy all application requirements. Recently, we have come across these issues in our project P6 “Scalable Computing Techniques” within the Research Training Group (RTG) 2450 - Tailored Scale-Bridging Approaches to Computational Nanoscience [6]. As scientists, we need a tool to facilitate the development of workflow applications for different workflow systems with minimum required knowledge of the underlying hardware and the technicalities of the workflow systems.

To this end, we have adopted a generative approach based on a simple abstract workflow description language (WConfig) to solve these issues effectively [7]. We have implemented the approach in the wfGenes tool [8] that is

available under an open source license. In this framework, the scientist describes the tasks (see Figure 1) and their data dependencies and then selects a target workflow system. With running wfGenes, the workflow graph of the application is constructed and the code for the target workflow systems is produced. Optionally, if a schema is provided by the target workflow system, the generated code is validated against the schema. All work steps within wfGenes shown in Figure 1 are processed automatically. The user only has to prepare the WConfig input and select the target system.

High-throughput and loosely coupled multiscale computing

There has been a request in RTG 2450 to translate an existing FireWorks workflow for computing the adsorption free energy in a solid – gas heterogeneous system to the SimStack workflow system [9]. We have not followed a translator approach but rather a generative approach that has led to the development of the first

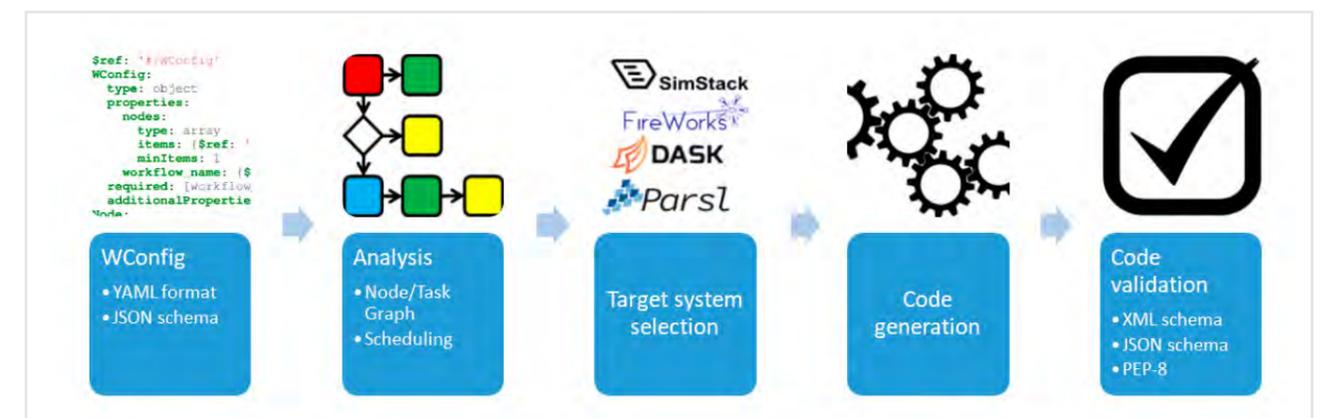


Figure 1: The work steps using wfGenes.

version of wfGenes. Indeed, the effort to design the application in WConfig from scratch and generate the input code for FireWorks and SimStack was negligible compared to a non-automated approach. After schema validation, to ensure valid functionality, the code generated by wfGenes has been verified by comparing to reference output data produced with the original FireWorks workflow [7].

Tightly coupled multiscale computing

Many applications in RTG 2450 include tightly coupled multiscale modelling and therefore require systems for task-based computing allowing scalable fine-grained parallelism. We have evaluated Dask [10] and Parsl [11], which are such systems for the Python language, and have found that they are helpful for composing such applications. However, these require

advanced programming skills in Python and high development effort especially for more complex applications, such as simulation of charge carrier transport in organic electronics [12]. This motivated us to extend wfGenes to generate Python code for these two systems from a declarative description in the same simple abstract workflow language (WConfig). This means that the developer does not have to program in Python at all. As a bonus, wfGenes generates code with optimal parallel execution using a static task graph scheduler (see Figure 1). For applications with a medium-size task graphs, with dozens of nodes, the current processing speed of wfGenes is sufficiently high to allow embedding wfGenes for interactive use in Jupyter notebooks (see Figure 2) or in workflow editors.

Generic tool

Although the first application use cases come from the materials science and nanoscience domains, the wfGenes tool is not limited to these applications. We expect that any workflow or task-based parallel application can be generated using wfGenes. Therefore, we believe that the tool can be adopted in other domains to provide users with a high-level method to implement task parallelism on personal workstations or supercomputers using one development environment. For example, complex data analyses, such that already rely on Dask, can directly benefit from wfGenes. Furthermore, wfGenes can be readily integrated to facilitate editing of complex models and analyses in the Jupyter-based Virtual Research Environment currently developed within the Joint Lab VMD (S. 38) [13].

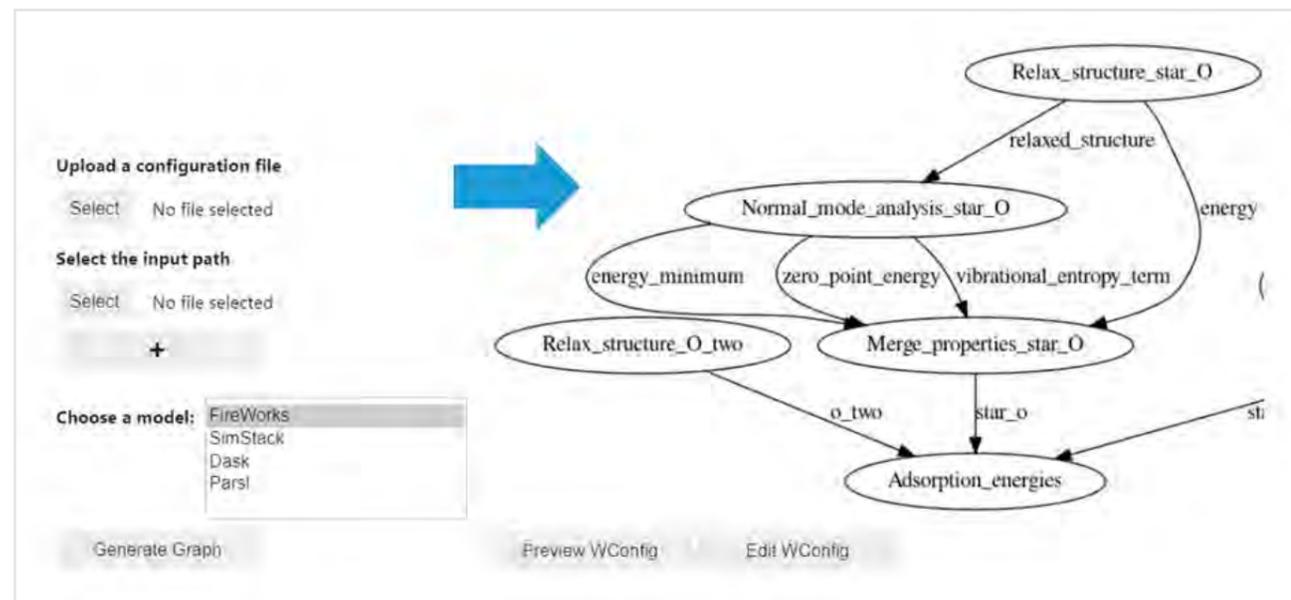


Figure 2: wfGenes's graphical user interface used in a Jupyter notebook

References

- [1] Computational Data Analysis Workflow Systems: in incomplete list, s.apache.org/existing-workflow-systems (accessed 2021-06-17)
- [2] UNICORE, www.unicore.eu/ (accessed 2021-06-17)
- [3] FireWorks materialsproject.github.io/fireworks/ (accessed 2021-06-17)
- [4] I. Kondov, P. Faubert, and C. Müller, Activity and Electrochemical Stability of a Chromium Modified Nickel Catalyst for Oxygen Reduction Reaction, *Electrochimica Acta* 236, 260 (2017). doi.org/10.1016/j.electacta.2017.03.123; P. Faubert, I. Kondov, D. Qazzazie, O. Yurchenko, and C. Müller, A Non-Noble Cr–Ni-Based Catalyst for the Oxygen Reduction Reaction in Alkaline Polymer Electrolyte Fuel Cells, *MRS Communications* 8, 160 (2018). doi.org/10.1557/mrc.2018.13
- [5] M. Vandichel, K. Laasonen, and I. Kondov, Oxygen Evolution and Reduction on Fe-Doped NiOOH: Influence of Solvent, Dopant Position and Reaction Mechanism, *Top Catal* 63, 833 (2020). doi.org/10.1007/s11244-020-01334-8; N. Akbari, I. Kondov, M. Vandichel, P. Aleshkevych, and M. M. Najafpour, Oxygen-Evolution Reaction by a Palladium Foil in the Presence of Iron, *Inorg. Chem.* 60, 5682 (2021). doi.org/10.1021/acs.inorgchem.0c03746
- [6] Research Training Group (RTG) 2450 Tailored Scale-Bridging Approaches to Computational Nanoscience, www.compnano.kit.edu/ (accessed 2021-06-17)
- [7] M. Roozmeh and I. Kondov, Workflow Generation with wfGenes, in *IEEE/ACM Workflows in Support of Large-Scale Science (WORKS)* (IEEE 2020), pp. 9–16. doi.org/10.1109/WORKS51914.2020.00007
- [8] wfGenes, git.scc.kit.edu/th7356/wfgenes (accessed 2021-06-17)
- [9] SimStack, www.simstack.de/ (accessed 2021-06-17)
- [10] Dask, dask.org/ (accessed 2021-06-17)
- [11] Parsl, parsl-project.org/ (accessed 2021-06-17)
- [12] P. Friederich, T. Strunk, W. Wenzel, and I. Kondov, Multiscale Simulation of Organic Electronics via Smart Scheduling of Quantum Mechanics Computations, *Procedia Computer Science* 80, 1244 (2016). doi.org/10.1016/j.procs.2016.05.495
- [13] Virtual Materials Design (VirtMat), www.materials.kit.edu/298.php (accessed 2021-06-17)

wfGenes: Produktivitätssteigerung bei der Entwicklung von Workflows

Im Vergleich zu anderen Anwendungsbereichen des wissenschaftlichen Rechnens bestehen für die rechnergestützte Materialforschung besondere Herausforderungen. Diese sind hauptsächlich auf die Vielzahl von Softwaremodulen zurückzuführen, die zur Durchführung einer Studie in eine neue Anwendung integriert werden müssen. Ein bestimmtes Softwaremodul erfüllt eine spezifische Aufgabe in der gesamten Simulation, die oft in Form eines wissenschaftlichen Workflows dargestellt wird. Um solche komplexen Simulationen auf Supercomputern zu steuern, werden spezialisierte Workflow-Managementsysteme eingesetzt. Aufgrund der unterschiedlichen Anwendungsanforderungen existiert derzeit eine große Vielfalt solcher Systeme. Dennoch erfordert der Einsatz dieser Systeme fortgeschrittene IT-Kenntnisse und einen erheblichen Aufwand für die Entwicklung, die Verifizierung und den Einsatz von Workflow-Anwendungen. Darüber hinaus ist die Übertragbarkeit auf

andere Anwendungen und Benutzergruppen aufgrund der verwendeten Workflow-Sprache eingeschränkt. Im Projekt wurde innerhalb des DFG Graduiertenkollegs 2450 zur Lösung dieser Probleme ein generativer Ansatz gewählt und in das Open-Source-Werkzeug wfGenes implementiert. Der Anwender beschreibt die einzelnen Anwendungsschritte und ihre Datenabhängigkeiten in einer deklarativen Sprache. Vollautomatisch prüft wfGenes die Eingaben, konstruiert den Workflow-Graphen. Zudem erzeugt und validiert wfGenes den Code für die gewählten Ziel-Workflow-Systeme. In der Praxis ermöglicht wfGenes den schnellen Entwurf von Anwendungen mit verschiedenen Anforderungen. wfGenes ist zudem ein generisches Werkzeug, das nicht nur auf die Materialforschung beschränkt ist. Zum Beispiel kann es für jede Task-Graph-basierte Modellierung und Datenanalyse eingesetzt werden.

Digitale Zwillinge aus dem virtuellen Materiallabor

Das Joint Lab „Virtual Materials Design“ im Helmholtz-Forschungsbereich „Information“ ist eine Kooperation zwischen den Helmholtz-Programmen „Engineering Digital Futures“, „Natural, Artificial and Cognitive Information Processing“ und „Materials Systems Engineering“ mit dem Hauptziel, die Materialentwicklung durch die Bereitstellung digitaler Zwillinge für Materialien, Bauteile und Geräte zu beschleunigen sowie eine virtuelle Forschungsumgebung zu entwickeln, die skalenübergreifende Simulationsmethoden und digitale Zwillinge zugänglich macht.

Ivan Kondov



JOINT LAB VIRTUAL MATERIALS DESIGN

Wissenschaftliche Workflows ermöglichen die Automatisierung komplexer Simulationen und Datenanalysen, speichern dauerhaft Herkunftsinformationen von Daten, die die Reproduzierbarkeit der Forschung verbessern, und ermöglichen das Auffinden und die Wiederverwendung der produzierten Daten. Die derzeit in Materialsimulationen am häufigsten verwendeten Workflow-Management-Systeme sind dennoch generisch und erfordern umfangreiche Programmier- und Kommandozeilen-Kenntnisse. Andererseits wird die Forschung in der rechnergestützten Materialwissenschaft derzeit mit Allzweck-Computersprachen wie C++, Fortran und Python durchgeführt, wobei domänenspezifische Komponenten über Programmierschnittstellen (API) verwendet werden. Die Verwendung einer domänenspezifischen Sprache anstelle von Allzwecksprachen kann die Produktivität der Wissenschaftlerinnen und Wissenschaftler bei der Verwaltung der Lebenszyklen von Computermodellen und Daten erheblich verbessern.

Im Projekt „Virtual Research Environment for Materials Modeling and Data Analysis“ (P16)¹ geht das SCC über diesen

Stand der Technik hinaus, indem es eine anwendungsfallbezogene domänenspezifische Sprache (DSL) entwickelt, welche die Objekte und die Methoden in den Domänen der Elektronenstruktur, Molekulardynamik und Phasenfeldmodellierung abdeckt. Die neu entwickelte DSL bietet auf die Anwendungsfälle zugeschnittene Ausdrücke und abstrahiert die wichtigen Merkmale von den zugrundeliegenden Workflow-Management-Systemen sowie von Modellierungs- und Analysewerkzeugen. Unterstützende Werkzeuge wie Parser, Interpreter und Validatoren werden die produktive Nutzung der DSL sicherstellen. Es ist zu erwarten, dass bestehende Modelle und Analysen in diesen drei Bereichen für eine breitere Nutzergemeinschaft aus dem Bereich der Materialforschung leicht zugänglich sein werden.

Basierend auf JupyterLab (SCC-News 2/2020, S. 14) werden die verfügbaren Rechenressourcen und Workflow-Management-Systeme nahtlos in eine interaktive virtuelle Forschungsumgebung (VRE) integriert und in einem dedizierten Repository etablierte Modelle und Analysen in Form von Python-Notebooks

bereitgestellt. In einer späteren Phase werden dann die DSL-unterstützten Werkzeuge in die neue Umgebung integriert und die Nutzung der DSL-Notebooks ermöglicht. Außerdem werden Übersetzungen der Modelle und Analysen im Repository basierend auf den ausgewählten Anwendungsfällen in DSL-Notebooks bereitgestellt. Alle Modelle und Analysen im Repository werden mit Hilfe von Continuous-Integration-Methoden und -Werkzeugen kontinuierlich gegen bestehende Ausgabedaten verifiziert und für den produktiven Einsatz innerhalb der VRE zur Verfügung gestellt.

Digital twins from the virtual laboratory

Scientific workflows enable automation of complex simulation and data analysis, persist provenance information improving reproducibility of research, and allow finding, analysis, and reuse of produced data. However, the currently most used workflow management systems are still difficult to use by materials scientists. Within the Joint Lab „Virtual Materials Design“ in the Helmholtz Research Field Information, the project team at SCC goes beyond these systems by developing a domain-specific language targeting a broader user community and facilitating the development of digital twins of materials. Furthermore, a Jupyter-based virtual research environment will be developed to provide easy-to-use tools for data analysis based on the new language and abstracting from the backend workflow systems.

¹ KIT-Webseite des Joint Lab VMD: www.materials.kit.edu/298.php

Abschluss des Pilotprojekts Helmholtz Analytics Framework

Im Pilotprojekt Helmholtz Analytics Framework entwickelten Forschende der Helmholtz-Gemeinschaft domänenspezifische Datenanalysetechniken und -algorithmen, erprobten diese und stellten sie der Data Science Community als generische Dienste und Werkzeuge für den allgemeinen Gebrauch zur Verfügung. In insgesamt acht Anwendungsfällen aus fünf Forschungsbereichen untersuchten sie wissenschaftliche Fragestellungen zu den Themen Erdsystemmodellierung, Neurologie, Luft- und Raumfahrt, Photonexperimenten und Strukturbiologie. Das Pilotprojekt wurde im März 2021 erfolgreich abgeschlossen.

Markus Götz, Alexander Schug

Oktober 2017, Jülich: offizieller Kick-off des Helmholtz Analytics Frameworks (HAF). Im Plenum wird intensiv über Datenformate, Analysemethoden und deren Implementierungen diskutiert und erste Möglichkeiten für ein Zusammenarbeiten sondiert. Das Pilotprojekt des Impuls- und Vernetzungsfonds (IVF) im Bereich Data Science läuft an und wird sich in den kommenden Jahren mit der Erforschung von Scientific Big Data Analytics beschäftigen. Als nicht mehr weg zu denkender Bestandteil der modernen Wissenschaft helfen diese Techniken bei der Beantwortung zentraler Forschungsfragen.

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

(Alexander Schug, Forschungsgruppe Multiscale Biomolecular Simulation, SCC, HAF-Mitglied WP1, Anwendungsfall Strukturbiologie)

Die Mitglieder des HAF-Projektes werden dafür zunächst domänenspezifische Datenanalysetechniken und -algorithmen entwickeln, dann in den Anwendungsfällen des Projektes erproben und abschließend als generische Dienste bzw. Werkzeuge für den allgemeinen Gebrauch verfügbar machen. Dabei kommt ein sogenannter Co-Design-Ansatz zum Einsatz. Das bedeutet, dass Wissenschaftlerinnen und Wissenschaftler aus den jeweiligen Domänen eng mit Expertinnen und Experten der Informationsverarbeitung zusammenarbeiten, um einen

effektiven und effizienten Entwurf und Gebrauch der Datenanalysemethoden zu ermöglichen. Hohe Datenkomplexität und rechenintensive Verfahren erfordern dabei den Einsatz wegweisender Technologien aus dem Datenmanagement und dem Hochleistungsrechnen, um Datenzugriff, -integration, -verarbeitung und -assimilation zu ermöglichen.

In insgesamt acht Anwendungsfällen aus fünf Forschungsbereichen werden wissenschaftliche Fragestellungen zu den Themen Erdsystemmodellierung, Neurologie, Luft- und Raumfahrt, Photonexperimenten und Strukturbiologie untersucht. Das Vorhaben wurde von der Helmholtz Gemeinschaft mit nahezu 6 Millionen Euro unterstützt. Insgesamt waren sechs Helmholtz Zentren – DESY, DKFZ, DLR, FZJ, HMGU, KIT – an dem Vorhaben beteiligt¹. Die Projektteilnehmer hatten ebenfalls Zugriff auf die Anfang 2017 gestartete Helmholtz Data Federation (HDF), einer föderierten Datenverarbeitungsinfrastruktur, die auch für die Erprobung der im HAF entwickelten Datenanalysemethoden und deren Implementierungen zur Verfügung stand.

März 2021, Webkonferenz: das letzte Zusammentreffen aller HAF-Projektmitglieder (Abbildung 1). Wieder wird viel diskutiert, dieses Mal dreht es sich aber vor allem um die Ergebnisse der Forschung und wie diese erreicht wurden. Alle Projektbeteiligten sind sich einig: Das Format des Projekts – nämlich neue Kooperationspartner, neue methodische Ansätze und die Möglichkeit, sich

von anderen Wissenschaftsdomänen inspirieren zu lassen – ist etwas ganz Besonderes gewesen. Das hat es so bisher nicht gegeben und das zeigt sich auch in den Ergebnissen. So haben die HAF'ler zum Beispiel mit Hilfe von Ansätzen des Maschinellen Lernens zu der Erforschung neuer Raketentreibstoffe [1], der Beschleunigung iterativer Gleichungslöser [2] oder auch der Nutzung von Kalmanfiltern für das Training künstlicher neuronaler Netze [3] beigetragen. Aber auch ohne die Nutzung künstlicher Intelligenzen konnten die Forschenden mehr über neurale Aktivitätsmuster in Gehirnen, die Hydrologie Europas oder den Entwurf von Tragflächen ziviler Flugzeuge lernen.

Der einzige Wermutstropfen: Die COVID-19-Pandemie. Sie hat den Forschungsgruppen zu schaffen gemacht, sowohl beim Finden neuer Kollegen und Kolleginnen als auch bei der Planung und

Das gesamte Projekt in Zahlen

- Insgesamt gab es 41 Veröffentlichungen – 24 Artikel in Fachzeitschriften, 6 Artikel in Konferenzbänden, 6 Konferenzvorträge, 4 Abschlussarbeiten und 1 Poster
- 26 Forschende
- 7 Projekttreffen
- 7 Open-Source-Softwareprojekte
- 1 Best-Paper-Award
- 1 Erna-Scheffler-Preis
- ... und 4-mal Nachwuchs – die nächste Wissenschaftlergeneration steht in den Startlöchern

¹ Deutsches Elektronen Synchrotron (DESY), Deutsches Krebsforschungszentrum (DKFZ), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Forschungszentrum Jülich (FZJ), Helmholtz Zentrum München (HMGU)



Abbildung 1: HAF All-hands Meeting, 25.03.2021, virtuell

Durchführung von Kooperationstreffen. Es ist allerdings schon ein Termin für ein alternatives Abschlussgrillen angepeilt.

Und im SCC?

Zwei große Themenkreise haben die Forschung im Steinbuch Centre for Computing geprägt. Zum einen die Frage, inwieweit man mehr über Biomoleküle lernen kann, und zum anderen, wie technische Werkzeuge für die Verarbeitung großskaliger, mehrdimensionaler Arraydaten aussehen können. Aber der Reihe nach:

Knapp zusammengefasst wollten wir mehr über die molekulare Grundlage des Lebens lernen, indem wir die bestmögliche Interpretation experimenteller Daten wie NMR-, CryoEM- und Kleinwinkelstreuendaten ermöglichen. Drei Promovierende haben in einer engen Kooperation mit dem FZJ und dem HMGU erfolgreich den Einsatz von High Performance Computing für solche strukturbio- logische Analysen vorangetrieben. Dabei wurde eine große Bandbreite von molekularbiologischen Systemen behandelt, die von

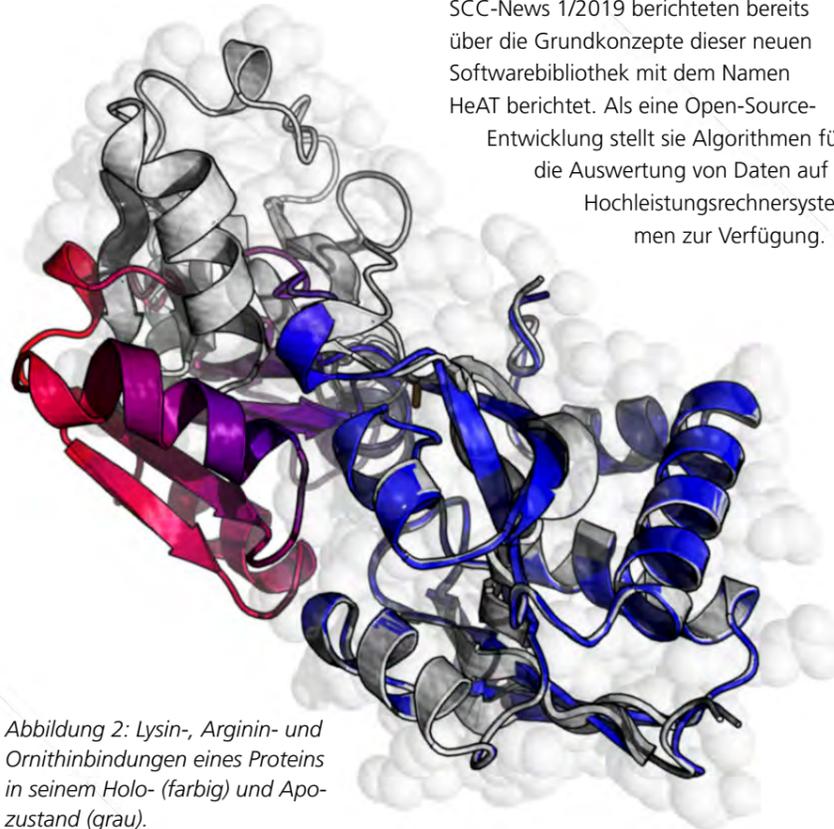


Abbildung 2: Lysin-, Arginin- und Ornithinbindungen eines Proteins in seinem Holo- (farbig) und Apo- zustand (grau).

kleinen Proteinen bis hin zu medizinisch relevanten Systemen wie Viren reichen. Der wissenschaftliche Erfolg wird durch erfolgreiche Veröffentlichungen in hochrangigen wissenschaftlichen Zeitschriften unterstrichen. So konnten Verfahren des maschinellen Lernens genutzt werden, um strukturbio- logische Information besser auszuwerten² (Abbildung 2). Besonders erfreulich war das Interesse der Öffentlichkeit an den wissenschaftlichen Inhalten zur rechnergestützten Strukturbio- logie. Die beteiligten Projektleiter vom SCC und FZJ waren gerngesehene Interviewpartner für Presse- und Medien- beiträge, die unter anderem im NDR, der Welt, der Süddeutschen und Frankfurter Allgemeinen Zeitung sowie in heise.de erschienen.

Drei andere Forschende haben sich dage- gen intensiv zusammen mit Kolleginnen und Kollegen aus dem FZJ und dem DLR mit der Bereitstellung eines Werkzeugs für die generische Verarbeitung dieser groß- skaligen Daten beschäftigt. Die SCC-News 1/2019 berichteten bereits über die Grundkonzepte dieser neuen Softwarebibliothek mit dem Namen HeAT berichtet. Als eine Open-Source- Entwicklung stellt sie Algorithmen für die Auswertung von Daten auf Hochleistungsrechnersyste- men zur Verfügung.

Die in der Programmiersprache Python geschriebene Software, erlaubt nicht nur die transparente Benutzung von Beschleunigerhardware wie Grafikkarten, sondern abstrahiert auch von explizitem Kommunikationscode durch den Pro- grammierer und gestattet so die Nutzung von Clustercomputern (Abbildung 3). Mittlerweile befindet sich HeAT in der produktiven Nutzung in diversen Helm- holtz-Zentren auch über die HAF-Partner hinaus. Seit Anfang dieses Jahres ist die Firma Intel Entwicklungspartner im HAF- Projekt.

„HAF war aus meiner Warte ein sehr erfolgreiches Projekt. Mir hat vor allem gefallen, wie sich Kooperationen nicht nur zwischen den Informa- tionsexperten und Anwen- dungsforschern gebildet ha- ben, sondern auch zwischen, teils sehr unterschiedlichen, Forschungsfeldern.“

(Björn Hagemeier, Juelich Supercomputing Centre (JSC), Forschungszentrum Jülich, HAF-Mitglied WP2)

Die Mitglieder des Helmholtz Analytics Frameworks blicken auf drei sehr erfolg- reiche Projektjahre zurück. Es konnten tolle Ergebnisse erreicht und viele Ant- worten auf Forschungsfragen gefunden werden. Die Zusammenarbeit im Koope- rationsverbund hat viel Spaß gemacht, und viele der Kooperationen werden auch nach Projektende erhalten bleiben. Ein ganz herzliches Dankeschön geht an alle Beteiligten.

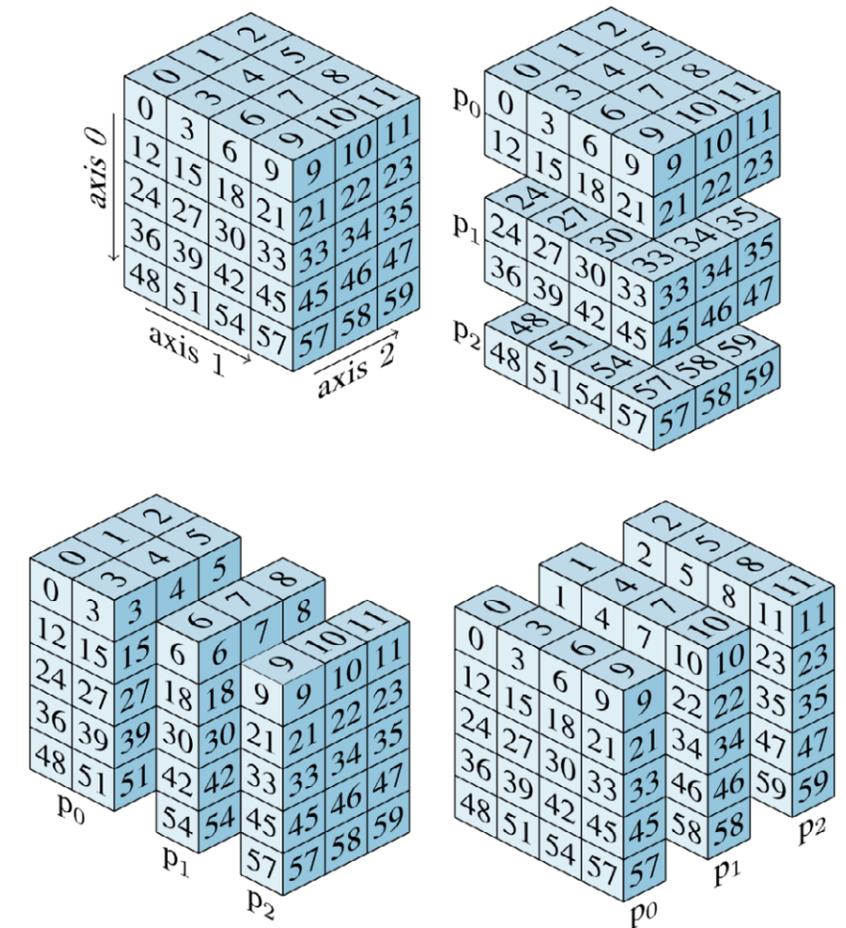


Abbildung 3: Daten-Dekomposition eine dreidimensionalen DNDarrays in HeAT.

Referenzen

- [1] Debus, C., Ruettgers, A., Petrarolo, A., Kobald, M. and Siggel, M., 2020. High-performance data analytics of hybrid rocket fuel combustion data using different machine learning approaches. In AIAA Scitech 2020 Forum (p. 1161).
- [2] Götz, M. and Anzt, H., 2018, November. Machine learning-aided numerical linear algebra: Convolutional neural networks for the efficient preconditioner generation. In 2018 IEEE/ACM 9th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems (scalA) (pp. 49-56). IEEE.
- [3] Yegenoglu, A., Krajsek, K., Pier, S. D., & Herty, M. (2020, July). Ensemble Kalman Filter Optimizing Deep Neural Networks: An Alternative Approach to Non-performing Gradient Descent. In International Conference on Machine Learning, Optimization, and Data Science (pp. 78-92). Springer, Cham.

Completion of the Helmholtz Analytics Framework

In the Helmholtz Analytics Framework pilot project, researchers from the Helmholtz Association developed domain-specific data analysis techniques and algorithms, tested them, and made them available to the data science community as generic services and tools for general use. In a total of eight use cases in five research areas, they investigated scientific issues related to earth system modeling, neurology, aerospace, photon experiments, and structural biology. The pilot project successfully completed in March 2021.

O3as: ein Dienst zur Beurteilung des Ozonzustands

Sachstandsberichte über stratosphärisches Ozon zu erstellen, ist eine wichtige Aufgabe für Klima- und Umweltstudien. Der O3as-Dienst stellt Wissenschaftlerinnen und Wissenschaftlern sowie allen am Ozonzustand Interessierten ein Werkzeug zur Verfügung, um eine Analyse von Ozondaten durchzuführen sowie Grafiken von hoher Qualität zu erzeugen und zuverlässig zu reproduzieren. Er ist einer der sogenannten „thematic services“ des EU Projekts EOSC-Synergy. Der Dienst nutzt Best Practices der Softwareentwicklung: Der Code ist Open-Source, gut gepflegt, öffentlich zugänglich und mit CI/CD-Methoden implementiert. Derzeit sind damit etwa 100 Klimamodel-Simulationen zugänglich. O3as wird im Herbst 2021 auf dem EOSC-Marketplace veröffentlicht werden.

Valentin Kozlov, Tobias Kerzenmacher¹

Die Überwachung von Ozon ist eine wichtige Aufgabe für Klima- und Umweltstudien. Der Abbau von Ozon in der stratosphärischen Ozonschicht führt dazu, dass mehr Ultraviolettstrahlung die Erdoberfläche erreicht und alle Lebensformen schädigt. Seit der Entdeckung des Ozonlochs in den achtziger Jahren erkannten Forschende auf der ganzen Welt die daraus resultierenden Gefahren. Mit der Verabschiedung des Montreal-Protokolls, das 1986 von den Vereinten Nationen ratifiziert wurde, kam es zu einem Verbot der Stoffe, die die Ozonschicht zerstören. Seitdem werden regelmäßig Berichte zur Bewertung des stratosphärischen Ozonabbaus erstellt [1]. Der letzte Bericht wurde 2018 veröffentlicht, der nächste ist für 2022 geplant. Der von SCC und dem Institut für Meteorologie und Klimaforschung – Atmosphärische Spurengase und Fernerkundung

(IMK-ASF) gemeinsam entwickelte Ozone Assessment Service (O3as) unterstützt Wissenschaftlerinnen und Wissenschaftler sowie alle Interessierten bei der Bestimmung von Ozontrends für verschiedene Teile der Welt. Er ist einer der „thematic services“ des Projekts EOSC-Synergy [2].

Verbesserung des Arbeitsablaufs

Für die Berichterstellung werden große Mengen von Klimamodellprojektionen analysiert, um eine Zeitreihe der Ozonkonzentration zu erstellen, Trends zu erkennen und den möglichen Zeitpunkt abzuschätzen, an dem das Ozon wieder das gleiche Niveau wie im Jahr 1980 erreicht. Der derzeitige Arbeitsablauf besteht aus mehreren manuellen Schritten, und Code ist nicht immer verfügbar. Das bedeutet, dass die Daten und Grafiken

zur zeitlichen Entwicklung des Ozons nicht einfach zu reproduzieren sind und Inkonsistenzen aufweisen, insbesondere bei der großen Vielfalt an möglichen Eingabeparametern.

Um diesen aufwändigen Workflow effizienter zu machen, bietet O3as [3] die Möglichkeit, die Daten benutzerfreundlich abzurufen und zu reproduzieren, sowie zuverlässig hochwertige Plots für die Wissenschaft zu erstellen. Dies beinhaltet auch die folgenden Best Practices der Softwareentwicklung: Die entwickelte Software ist Open-Source, gut gepflegt und öffentlich zugänglich. Sie wird mit sogenannten Continuous Integration- und Continuous Delivery-Pipelines (CI/CD) auf Basis von Jenkins implementiert. Dies ermöglicht auch automatisierte Code-Tests. Die Komponenten des O3as-Dienstes werden in Form von Docker-Images

ausgeliefert, der Code sowie der Dienst sind gut dokumentiert.

Technische Dienstarchitektur

Der Dienst O3as besteht aus mehreren Komponenten. Um einen komfortablen Dienst anzubieten, werden zunächst die Original-Rohdaten von verschiedenen Ozonmodellen heruntergeladen und in der Large Scale Data Facility (LSDF) gespeichert (Abbildung 1). Diese Daten werden im Umfang von mehreren zehn Terabyte werden mit Hilfe der o3skim-Komponente auf die notwendigen Parameter reduziert. Dieser Prozess läuft regelmäßig auf einem der SCC HPC-Systeme (bwUni-Cluster bzw. HoreKa) und speichert die reduzierten Daten in der LSDF. Nutzende, die Ozondaten auswerten möchten, können zunächst die Webanwendung, o3webapp, verwenden und die Anfrage konfigurieren. Diese Anfrage wird an die o3api weitergeleitet, die die reduzierten Daten lokalisiert, sie entsprechend der Anfrage auswählt, verarbeitet und in wenigen Sekunden einen qualitativ hochwertigen Plot (Abbildung 2) liefert. Wissenschaftler und Wissenschaftlerinnen erhalten nach einer föderierten Authentifizierung (z. B. über EGI Check-in und DFN AAI) weitere Funktionen vom Dienst. Ferner kann man auch Daten und Plots abrufen, indem man direkt mit der o3api kommuniziert.

O3skim vereinheitlicht die ursprünglichen Rohdaten, da diese je nach Institution, die sie berechnet hat, in verschiedenen Formaten vorliegen können. Außerdem reduziert o3skim die vierdimensionalen Datensätze auf zwei Dimensionen, indem es die Längengrade mittelt und die Höhen integriert. Dieser Schritt ist rechenintensiv und erfordert leistungsfähige HPC-Ressourcen. Eine „dockerisierte“ Lösung sowie das Tool udocker [4] sorgen dafür, dass diese Komponente einfach auf verschiedene HPC-Systeme übertragen und dort betrieben werden kann.

O3api basiert auf der OpenAPI-Spezifikation, Swagger, Flask und Connexion. Sie läuft im Kubernetes-System, das bei einem der EOSC-Provider im Rahmen des Projekts EOSC-Synergy bereitgestellt wird. Eine erste Version der o3webapp wurde von einem studentischen Team im Rahmen des Praktikums für Software Entwicklung (PSE) in der Informatik entwickelt und soll weiter verbessert werden. Alle Komponenten nutzen die Empfehlungen zur Software-Qualitätssicherung, die von EOSC-Synergy entwickelt und gefördert werden, und nutzen insbesondere die JePL-Bibliothek (Jenkins Pipeline Library) zum Aufbau von CI/CD-Pipelines. Alle Dienste werden mit Hilfe des Tools Uppti-me überwacht.

Ergebnisse und Ausblick

Der Dienst O3as ermöglicht heute schon den Zugang zu etwa 100 Modellsimulationen der Chemistry-Climate Model Initiative (CCMI-1) sowie zu zwei auf Beobachtungen basierenden Datensätzen: SBUV Merged Ozone Dataset (MOD)

Referenzen

- [1] [csl.noaa.gov/assessments/ozone/](https://www.noaa.gov/assessments/ozone/)
- [2] www.eosc-synergy.eu/
- [3] o3as.data.kit.edu/
- [4] J.Gomes et al, Computer Physics Communications, 232, p.84, 2018,

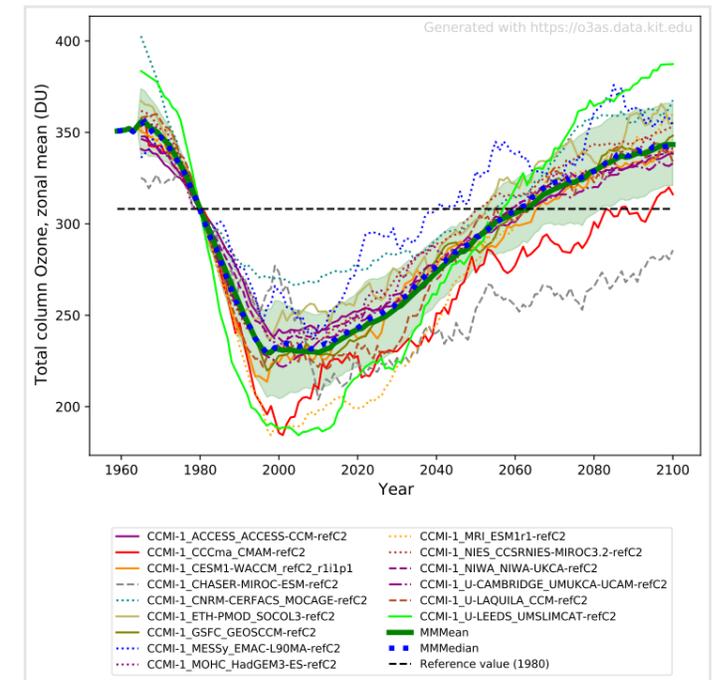


Abbildung 2: Beispielabbildung für die vom Dienst O3as erzeugte Ozongesamtsäule (Total Column Ozone).

der NASA und ESACCI des DLR, einem Datensatz aus der Climate Change Initiative der ESA, bereitgestellt vom DLR. Der Dienst soll bis Herbst 2021 zusammen mit anderen thematischen Services aus dem Projekt EOSC-Synergy auf dem EOSC-Marketplace veröffentlicht werden.

O3as: A service for assessing the state of the ozone layer

Ozone assessment is an important task for Climate and Environmental studies. The O3as service is providing a tool for scientists and everyone interested in the ozone assessment to establish the analysis of ozone data as well as to reliably generate and reproduce high quality figures. The service implies the best practices of software development: the code is open-source, well maintained, publicly accessible and implemented with CI/CD. Currently about 100 climate model simulations are accessible via the service and it will be published at the EOSC-Marketplace in autumn 2021.

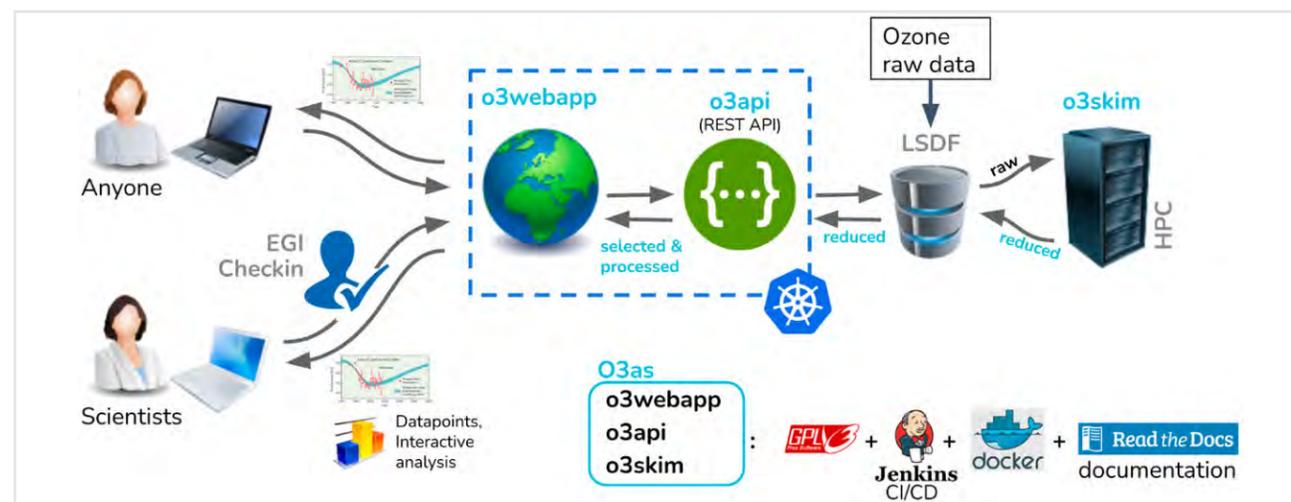


Abbildung 1: O3as Komponenten und die Benutzerinteraktion mit dem O3as-Dienst.

Research Software Engineering in der Lehre

Nachhaltige Entwicklung von Forschungssoftware, Teil des Research Software Engineerings – kurz RSE, wird immer wichtiger, und Studierende müssen geschickt an die damit verbundenen Konzepte herangeführt werden. Die am SCC forschende Gruppe Fixed-Point Methods for Numerics at Exascale (FiNE) hat sich auf die Entwicklung nachhaltiger Software für zukünftige Supercomputer spezialisiert und ein Konzept für die Lehre erarbeitet.

Hartwig Anzt

Die Lehrveranstaltung „Numerical Linear Algebra for Scientific High Performance Computing“ der Forschungsgruppe FiNE beinhaltet weit mehr als auf Folien präsentierte Lehrinhalte.

Tatsächlich verfolgt die Gruppe ein Lehrkonzept, das die Studierenden mit einer gemeinsamen Studienarbeit so an das Thema heranhöhrt, als wären sie Teil eines realen, kollaborativen Forschungssoftwareprojekts. Genauer gesagt bearbeiten die Studierenden Hausaufgaben, die neben einem theoretischen Teil auch einen Coding-Teil beinhalten, der als nachhaltige Forschungssoftware realisiert werden muss. Dies beinhaltet Aspekte wie Code-Dokumentation, Unit-Tests und Continuous Integration (Abbildung 1). Die entwickelten Lösungen werden dann als Pull-Request an ein zentrales Softwarerepository (Git-Instanz) gesendet. Genau wie in einem realen Forschungssoftware-Entwicklungsprozess übernehmen die studentischen „Research Software Engineers“ die Rolle von Kolleginnen und Kollegen und überprüfen den Code anderer im Round-Robin-Verfahren.

Nach einer Review-Phase (Abbildung 2) werden die Studierenden ermutigt, ihre Lösungen anhand des Feedbacks zu verbessern. Die Hausaufgabe wird durch das Zusammenführen der Pull-Requests abgeschlossen.

Dieser Workflow ermöglicht es den Studierenden nicht nur, die relevanten Studieninhalte kennenzulernen, sondern auch den Workflow einer nachhaltigen Forschungssoftware-Entwicklung in einem kollaborativen Umfeld zu erleben, wie das in der Wissenschaft gang und gäbe ist. Somit gewinnen die Studierenden an Expertise und Erfahrung im Einsatz von Continuous Integration, Unit Testing und Clean Software Development.

Ergänzt durch die technischen Studieninhalte und Aufgabenstellungen zu Algorithmen und Programmiermodellen für paralleles Rechnen auf Multicore- und Manycore-Architekturen, verteiltes Rechnen und fortgeschrittene numerische lineare Algebra-Methoden, erweist sich der Kurs als effektives Modul, das die Studierenden auf das wissenschaftliche

Hochleistungsrechnen vorbereitet. Die Benotung berücksichtigt alle Aspekte der Kursteilnahme: Die Leistung in einer Projektarbeit, die Leistung bei den theoretischen Hausaufgaben, die Korrektheit und Qualität der Codieraufgaben, die Qualität des Feedbacks in der Review-Phase und die Einarbeitung von Verbesserungsvorschlägen aus der Begutachtung.

Dieses Lehrkonzept umzusetzen, bedeutet für die Forschungsgruppe FiNE einen erheblichen Arbeitsaufwand. Die beiden Mitarbeiter, die den Hausaufgaben-Workflow betreuen, sehen dennoch, dass sich der Aufwand auszahlt: „So bilden wir unsere eigenen Mitarbeiter aus und bereiten sie mit den Fähigkeiten und der Erfahrung vor, die wir brauchen.“ In einem wissenschaftlichen Beitrag [1], der auf der International Conference on Computational Science (ICCS) vorgestellt wurde, wird das Lehrkonzept und der Hausaufgaben-Workflow ausführlich beschrieben. Außerdem werden die Rückmeldungen der Kursteilnehmer bezüglich des Arbeitsaufwands, des Schwierigkeitsgrads der Übungen und des reibungslosen Ablaufs der Übungen zusammengefasst.

Obwohl die ICCS in diesem Jahr virtuell statt fand, stieß die Vorstellung des Konzepts auf großes Interesse, und die Forschungsgruppe FiNE ist zuversichtlich, dass weitere Institutionen diesem Weg folgen werden.

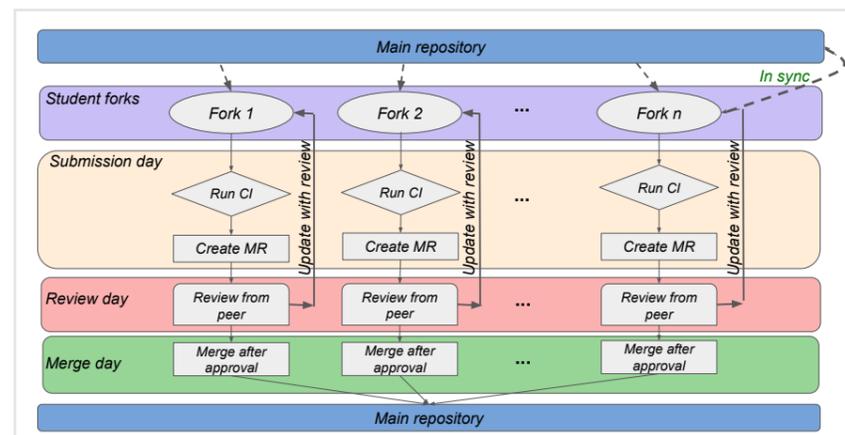


Abbildung 1: Workflow der Studierenden bei der Bearbeitung der Coding-Hausaufgaben.

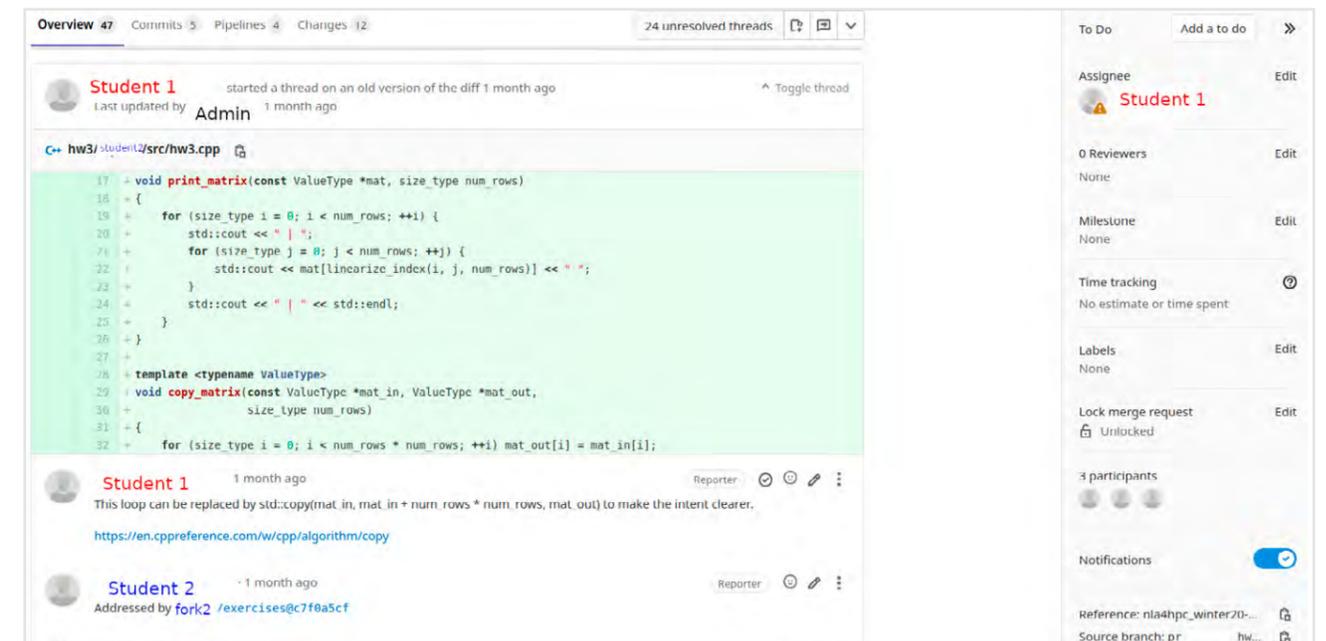


Abbildung 2: Studierende kommentieren und verbessern die Coding-Hausaufgaben der Kolleginnen und Kollegen.

Research Software Engineering in teaching

The course “Numerical Linear Algebra for Scientific High Performance Computing” offered by the research group FiNE involves a lot more than classical lecturing. In fact, the group pursues a teaching concept that involves the students in the coursework as if they were part of a collaborative research software project. In more detail, the students committing to the course need to complete homework assignments that include both a theoretical part and a coding part that has to be realized as sustainable research software, including aspects like code documentation, unit tests, and continuous integration. The developed solutions are then submitted to a central git instance as a pull request, like in a research software development process, the students take the role of colleagues and peer-review others' code in a round-robin fashion. After a review phase, the students are encouraged to improve their solutions according to the feedback, before the homework assignment is completed by merging the pull requests. This workflow allows the students to not only learn about the course content, but also experience the workflow of sustainable research software development in a collaborative setting. Obviously, the students complete the course with comprehensive expertise and experience in using continuous integration, unit testing, and clean software development.

Complemented by the technical course content and assignments covering algorithms and programming models for parallel computing on multicore and manycore architectures,

distributed computing, and advanced numerical linear algebra methods, the course proves to be an effective module preparing students for scientific high performance computing. The grading of the students accounts for all aspects of the course participation: the performance in a class project, the performance in the theoretical homework assignments, the correctness and quality of the coding assignments, the quality of feedback provided in the review phase, and the incorporation of improvement suggestions received from the reviewers.

This teaching concept does come with a significant workload for the FiNE research group. However, the two FiNE group members handling the homework assignment workflow still assess the concept as paying off: “This way we educate our own employees and prepare them with the skills and experience we need.” In a scientific paper that was presented at the ICCS conference, they describe the teaching concept and the homework assignment workflow in detail. They also summarize the feedback they received from the course participants with respect to the workload, the toughness of the exercises, and the smoothness of the exercise workflow. Despite ICCS being virtual this year, the presentation of the concept has attracted quite some interest, and the research group FiNE is positive that more institutions will follow this path.

[1] P. Nayak et al. “A collaborative peer review process for grading coding assignments in coursework” in Computational Science – ICCS 2021 21st International Conference, Krakow, Poland, June 16–18, 2021, Proceedings, Part VI.

Foto-Impressionen der HoreKa-Einweihung

Der neue Supercomputer des KIT zählt zu den 15 schnellsten Rechnern in Europa und zu den 13 energieeffizientesten der Welt. Die enorme Rechenleistung des Hochleistungsrechners Karlsruhe, kurz HoreKa, macht neue Erkenntnisse etwa in Klima-, Energie- und Medizinforschung möglich.

Am 30. Juli 2021 hat die baden-württembergische Ministerin für Wissenschaft, Forschung und Kunst Theresia Bauer die 15-Millionen-Euro-Maschine der Wissenschaft übergeben.



Feierliche Inbetriebnahme des Hochleistungsrechners Karlsruhe durch (v.l.) Prof. Dr. Corinna Hoose (Professorin für Theoretische Meteorologie), Prof. Dr. Martin Frank (SCC-Direktor), Theresia Bauer (Ministerin für Wissenschaft, Forschung und Kunst Baden-Württemberg), Prof. Dr. Oliver Kraft (Vizepräsident für Forschung des KIT).



40 °C/45 °C: Ministerin Bauer fühlt den Temperaturunterschied des Kühlwassers bei Ein- und Austritt in den Cluster.



Führung durch den Kaltgang des Supercomputers HoreKa (v.l.n.r. Theresia Bauer, Oliver Kraft, Corinna Hoose, Martin Frank, Simon Raffener). Ein extrem guter Luftaustausch im Kaltgang ermöglicht es, hier die Führung auch mit Abständen kleiner als 1,5m durchzuführen.



Simon Raffener vom SCC zeigt Ministerin Bauer und Vizepräsident Kraft das Innenleben eines HoreKa-Rechenknotens.

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 (verantwortlich),
Karin Schäufele, Andreas Ley
E-Mail: redaktion@scs.kit.edu

Gestaltung, Satz und Layout
Heike Gerstner, Hella Grolmus
AServ – Crossmedia – Grafik
Karlsruher Institut für Technologie (KIT)
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen

Titelfoto
Ministerin Theresia Bauer und Vize-
präsident Oliver Kraft bei der Einwei-
hung des Supercomputers HoreKa
am KIT. (Foto: Riccardo Prevete)

Fotos
SCC, KIT, Riccardo Prevete

Druck
Systemedia GmbH, 75449 Wurmberg

Erscheinungstermin dieser Ausgabe
August 2021

www.scc.kit.edu/publikationen/scc-news

Der Nachdruck und die elektronische
Weiterverwendung sowie die Weiter-
gabe 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
www.scc.kit.edu/twitter
contact@scc.kit.edu