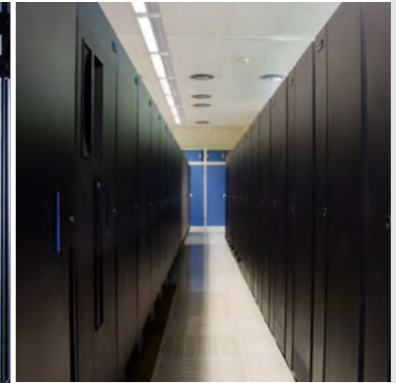


# SCC *news*

Steinbuch Centre for Computing

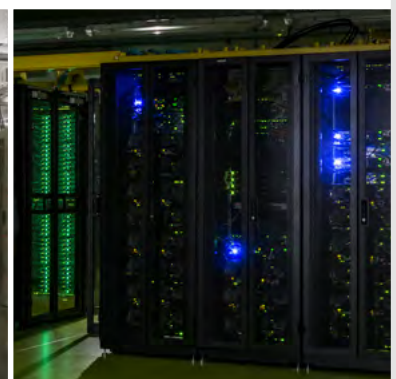
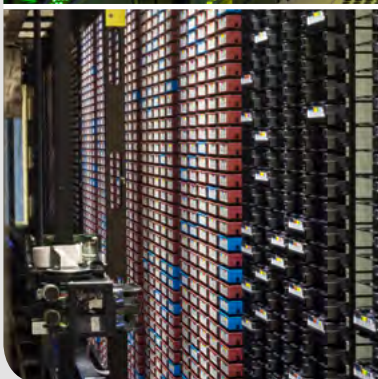
Ausgabe 2 | 2016



**Ein offener Hafen für Forschungsdaten**  
*An Open Haven for Research Data*

**Landesdienst bwDataArchiv geht in Betrieb**  
*State-wide Service bwDataArchiv goes into Operation*

**Smart Data Innovation Conference 2016 am SCC**  
*Smart Data Innovation Conference 2016 at SCC*



## *Liebe Leserinnen und Leser,*

im Sommer dieses Jahres hat der von Bund und Ländern berufene Rat für Informationsinfrastrukturen (Rfll) Empfehlungen zum Forschungsdatenmanagement in Deutschland veröffentlicht. Darin macht der Rat den Vorschlag eine „Nationale Forschungsdateninfrastruktur“ als Netzwerk mit übergreifenden Zentren zu errichten. Durch deren ausgeprägte Expertise in der Beschreibung, der Verarbeitung, und Speicherung bis hin zur Langzeitarchivierung von Forschungsdaten, sollen tragfähige und nachhaltige Strukturen im Forschungsdatenmanagement geschaffen werden.

Nur etwa drei Monate später gab die Helmholtz-Gemeinschaft bekannt, mit Beginn 2017 ein solches Netzwerk aus multi-thematischen Datenzentren – die Helmholtz Data Federation (HDF) aufzubauen. Die Helmholtz-Zentren betonen dabei ihre Bereitschaft, die notwendigen Dateninfrastrukturen zu schaffen, dort Forschungsdaten langfristig zu bewahren und sie anderen zur Nachnutzung so offen wie möglich zugänglich zu machen (S. 18).

Das SCC am KIT koordiniert die HDF-Aktivitäten und bringt hier eine über Jahre erlangte Erfahrung in der Handhabung von großen Mengen wissenschaftlicher Daten ein. Ein wichtiger Baustein in diesem System ist der gerade in Betrieb gegangene Service bwDataArchiv, der von Wissenschaftseinrichtungen des Landes zur nachhaltigen Speicherung und Archivierung von Forschungsdaten genutzt werden kann (S. 6).

Neben den großen Datenmengen aus der Wissenschaft, die am SCC gespeichert und archiviert werden, bietet das SCC auch Dienste für die Smart Data-Forschung an. Das am SCC ansässige Smart Data Innovation Lab hat nun ein Projektjahr erfolgreich in Kooperation mit Wirtschaftsunternehmen gemeistert. Die Ergebnisse der Projekte wurden auf der ersten Smart Data Innovation Conference am KIT vorgestellt und diskutiert (S. 4).

Noch etwas in eigener Sache: Unser Kollege Hannes Hartenstein lässt sein Amt im Direktorium des SCC seit Oktober diesen Jahres ruhen, um sich verstärkt der Fortentwicklung der Governance der Informationsverarbeitung und -versorgung am KIT zuwenden zu können. Wir wünschen ihm viel Erfolg und freuen uns auf die weitere gute Zusammenarbeit.

Viel Freude beim Lesen.

*Bernhard Neumair, Achim Streit*



## *Dear reader,*

In the summer of this year, the Council for Information Infrastructures (Rfll) published recommendations on research data management in Germany. In the recommendations, the Council proposes a National Research Data Infrastructure conceived as a network with overarching centers. Their extensive expertise in the description, processing, storage and long-term archiving of research data enables establishment of a sustainable research data management structures.

Only about three months later, the Helmholtz Association announced the development of such a network of multi-thematic data centers - the Helmholtz Data Federation (HDF). With the HDF, the Helmholtz Centers emphasize their readiness to create the necessary data infrastructure, to maintain research data in the long term for future use and to make the results as open as possible (page 18). The SCC of KIT coordinates the activities of the HDF and provides its experience gained over many years in the handling of large quantities of scientific data. An important building block in this system is the bwDataArchiv service that has just been put into operation. The service can be used by scientific institutions of the state of Baden-Wuerttemberg and offers sustainable long term storage and archiving of research data (page 6).

In addition to the large amounts of data collected and archived, the SCC also delivers services for smart data research. The Smart Data Innovation Lab conducted a successful project year with increased cooperation with business enterprises. The results were presented and discussed among participants at the first Smart Data Innovation Conference at KIT (page 4).

On a final note: Our colleague Hannes Hartenstein temporary left the Board of Directors of the SCC since October in order to dedicate more time to the further development of the governance of information, communication and information technologies at KIT. We wish him lots of success and look forward to further cooperation.

Enjoy reading

*Bernhard Neumair, Achim Streit*

# Inhaltsverzeichnis

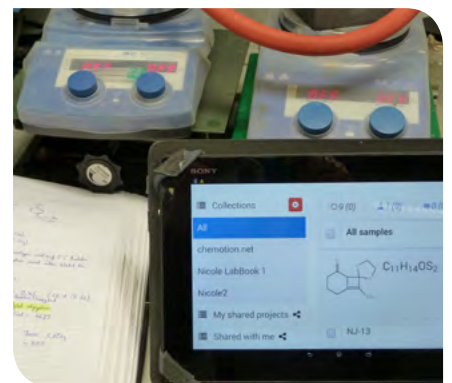
## DIENSTE UND INNOVATION

- 04 Smart Data Innovation Conference 2016 am SCC
- 06 Landesdienst bwDataArchiv geht in Betrieb
- 08 Virtuelle Graphikworkstations im Betrieb – ein Projektfazit
- 11 ILIAS 5 – Update des E-Learning-Systems des KIT



## FORSCHUNG UND PROJEKTE

- 13 V-FOR-WaTer – Eine virtuelle Forschungsumgebung für die Wasser- und terrestrische Umweltforschung
- 14 EUDAT Operations Management
- 16 Gründung der EUDAT Collaborative Data Infrastructure
- 17 bwDIM – Vereinfachter Datenfluss im Forschungsdatenmanagement
- 18 Ein offener Hafen für Forschungsdaten
- 18 Abgeschlossene Promotionsarbeit
- 19 Laborbuch 2.0 – digital und dezentral verfügbar
- 20 Sternenstaub, Wasserflöhe und Hochleistungsrechner:  
3. bwHPC-Symposium am Universitätsrechenzentrum Heidelberg



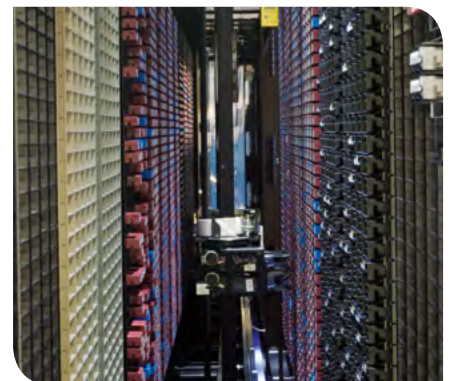
## STUDIUM UND WISSENSVERMITTLUNG

- 22 „Erste Hilfe“ für Erstsemester
- 23 Internationale GridKa Schule – Data Science on Modern Architectures
- 24 Oktober 2016 – der Cyber Security Month



## VERSCHIEDENES

- 25 Schülerkolloquium Simulierte Welten am SCC
- 26 31. International Supercomputing Conference
- 27 Forschungsdaten managen – E-Science-Tage 2017
- 27 Inside SCC – hohes Medien- und Besucherinteresse in 2016
- 27 Impressum & Kontakte



## Smart Data Innovation Conference 2016 am SCC

Das SCC hat im Oktober 2016 die erste Smart Data Innovation Conference organisiert. Im Rahmen dieser Konferenz wurden die Ergebnisse aus dem ersten Projektjahr des Smart Data Innovation Labs öffentlich vorgestellt und mit dem angereichten Fachpublikum diskutiert.

*Nico Schlitter*

Das Smart Data Innovation Lab (SDIL) ist ein seit 2014 bestehendes bundesweites Labor für Smart Data-Forschung auf Basis von In-Memory Big Data-Technologien. SDIL ist als Austausch- und Betriebsplattform konzipiert, um die Kooperation zwischen Wirtschaft, und Forschung im Bereich Big Data- und Smart Data-Technologien zu beschleunigen. Unternehmen und Forschungseinrichtungen haben sich dazu als Partner in den vier SDIL-Communities Industrie 4.0, Energie, Smart Cities und personalisierte Medizin organisiert,

um realitätsnahe Forschung in diesen strategisch wichtigen Feldern zu fördern.

Die Hauptkomponenten der SDIL-Plattform umfassen die In-Memory Datenbanken HANA von SAP und Terracotta der Software AG, sowie die Big Data-Lösungen der IBM Watson Foundation, inkl. SPSS Modeler und BigInsights. Die notwendigen Hardwareressourcen wurden von den Projektpartnern als Leihstellung ins Smart Data Innovation Lab eingebracht und werden vom SCC betrieben

(siehe auch SCC-News 2015, S. 4 und lookIT Ausgabe 3/2016, S. 30).

Die Projektmitarbeiter am Karlsruher Institut für Technologie (KIT), am Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) und am Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme (IAIS) untersuchen wissenschaftliche Fragestellungen aus den Bereichen Dateninfrastruktur, -analyse, -integration, -kuratierung und -anonymisierung.

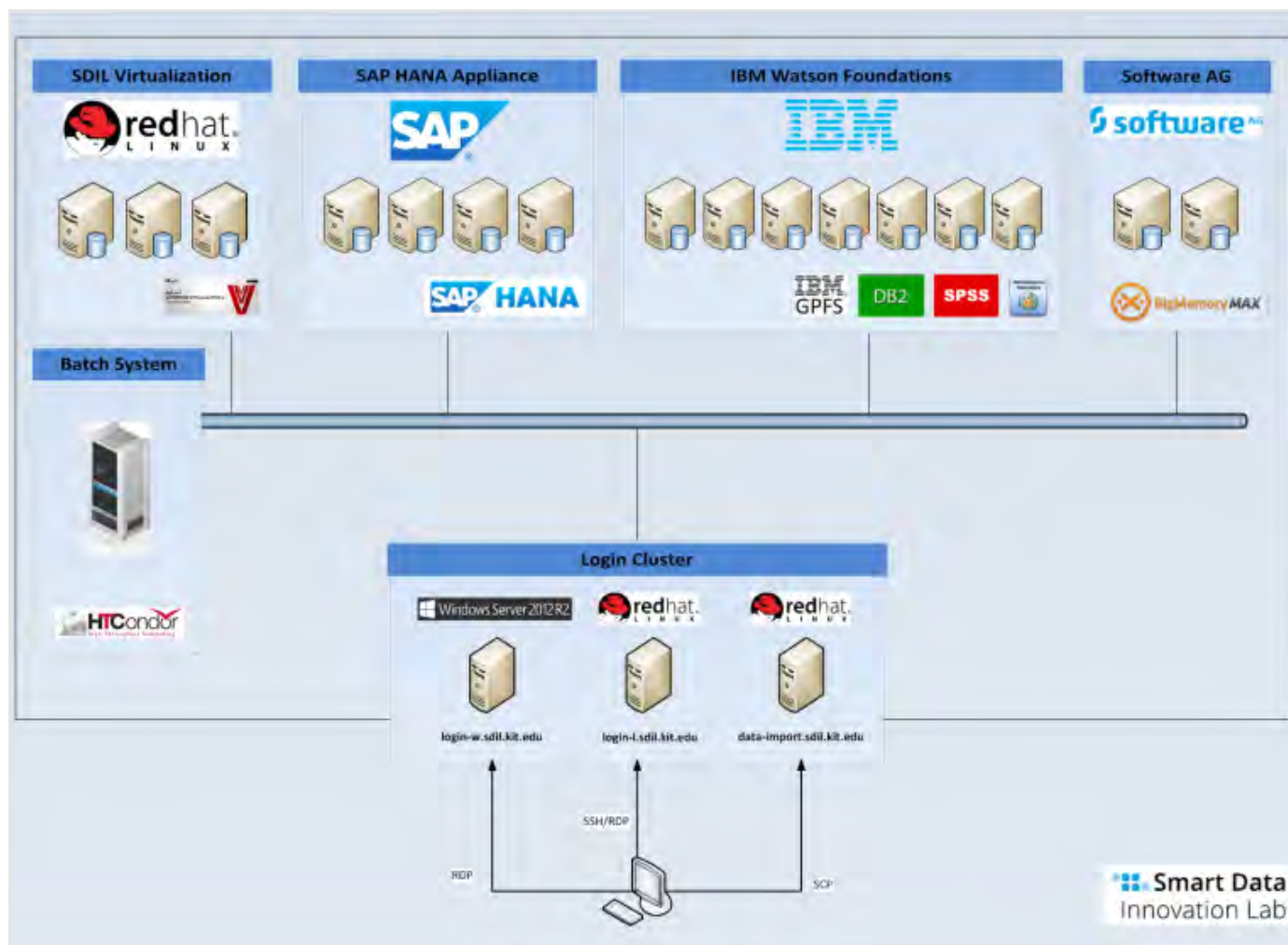


Abb. 1: Schematische Darstellung der SDIL-Plattform

Die Ergebnisse aus dem ersten Projektjahr des Smart Data Innovation Labs wurden am 12. und 13. Oktober 2016 im Rahmen der Smart Data Innovation Conference (SDIC) öffentlich vorgestellt und mit dem angereisten Fachpublikum diskutiert.

Der erste Konferenztag war durch angeregte Diskussionen in den Innovation Communities geprägt, welche das Ziel hatten neue Projektideen zu entwickeln und entsprechende Forschungsk Kooperationen zu initiieren. Am Folgetag wurden, im Rahmen von wissenschaftlichen Vorträgen, die Projektergebnisse der Projektpartner präsentiert. Eine Postersession rundete das wissenschaftliche Programm des Vormittags ab.

Der Nachmittag war der praktischen Einführung in die vielfältigen Möglichkeiten der SDIL-Plattform gewidmet. In drei parallel stattfindenden Tutorials konnten die Teilnehmer, unter Anleitung des SDIL Support Teams, die Plattform hautnah kennenlernen und testen:



Abb. 2: Wissenschaftlicher Fachvortrag auf der SDIC über Big Data Analysen

Morris Riedel, vom Forschungszentrum Jülich, gab eine praktische Einführung in die Datenanalyse und demonstrierte wie typische Klassifikationsprobleme mit Hilfe von parallelisierten Support-Vektor-Maschinen auf dem mit htCondor und MP betriebenen SDIL-Cluster gelöst werden können.

Michael Vössing, vom Karlsruhe Service Research Center (KSRI) am KIT, konnte den Teilnehmern näher bringen, wie die umfangreichen, im IBM SPSS Modeler enthaltenen Analyseverfahren durch die Nutzung des IBM Watson Clusters beschleunigt werden können.

Die Nutzung der SAP HANA Datenbank wurde von Markus Fath (SAP Deutschland) demonstriert. Besonderen Eindruck hinterließen dabei die Funktionalitäten, die über die typischen Funktionen eines klassischen relationalen Datenbanksystems hinausgehen.

Aufgrund des durchweg positiven Feedbacks der Konferenzteilnehmer haben wir beschlossen, die Smart Data Innovation Conference auch im nächsten Jahr am Steinbuch Centre for Computing durchzuführen.



Abb. 3: Projektleiter Nico Schlitter erläutert die Komponenten des Smart Data Innovation Labs.

### Smart Data Innovation Conference 2016 at SCC

The Smart Data Innovation Lab (SDIL) offers big data researchers access to a large variety of big data and in-memory technologies. Industry and science collaborate closely to find hidden value in big data and generate smart data. In October 2016, SCC hosted the Smart Data Innovation Conference (SDIC2016) and presented first results of the SDIL research project to the public. In addition, the SDIC2016 participants had the opportunity to participate in tutorials and hands-on trainings and get familiar with the unique potential of the SDIL platform.

## Landesdienst bwDataArchiv geht in Betrieb

bwDataArchiv, der neue Landesdienst für die Langzeitspeicherung und Archivierung von Forschungsdaten, geht in Betrieb. Mit dem neuen Dienst lassen sich große wissenschaftliche Datensätze über 10 Jahre und mehr zuverlässig speichern.

Bettina Bauer, Jos van Wezel

Das Landesprojekt bwDataArchiv schließt zum Jahresende seine dreijährige Projektstätigkeit ab und startet nun mit dem gleichnamigen Dienst. bwDataArchiv ermöglicht es, Wissenschaftlerinnen und Wissenschaftlern in Baden-Württemberg, Forschungsdaten zu archivieren. Erste Pilotnutzer speichern Daten, z. T. große Mengen, schon seit Mitte 2015 erfolgreich mit dieser Technologie. Dabei handelt es sich vor allem, aber nicht nur, um Nutzer des Projektpartners HLRS<sup>1</sup>.

Zeitgleich mit dem Projekt bwDataArchiv und ebenso mit Beteiligung des SCC, wurden weitere Langzeitspeicherdienste für Forschungsdaten aufgebaut, die die Speicherinfrastruktur von bwDataArchiv nutzen, jedoch über einen eigenen Nutzerkreis für Spezialanwendungen bedienen und dafür eine eigene Nutzerverwaltung betreiben. bwDataArchiv arbeitet mit den Diensten RADAR<sup>2</sup> und bwDataDiss<sup>3</sup> zusammen. Eine internationale Ausrichtung der Aktivitäten erfolgt

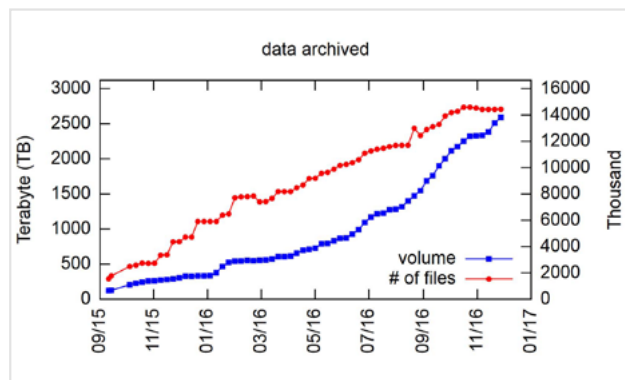


Abb. 1: Gespeicherte Datenmengen und Anzahl an Dateien seit Inbetriebnahme des HPSS-Speichersystems

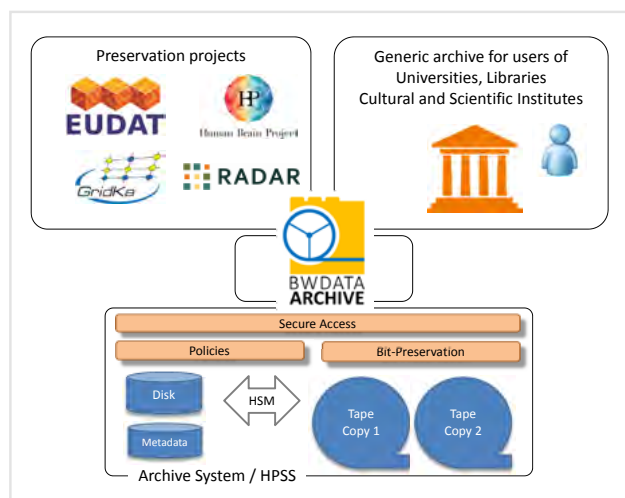


Abb. 2: bwDataArchiv bietet Wissenschaftlern von Einrichtungen in Baden-Württemberg sowie unterschiedlichen Forschungsprojekten mit hohem Datenaufkommen einen verlässlichen Langzeitspeicher.

durch die Einbindung des EU-Projekts EUDAT, das ihren Nutzern den Dienst B2SAFE zur Verfügung stellt. Zusammen mit dem Dienst B2FIND entsteht so ein Archivspeicherangebot für internationale Communities. EUDAT beteiligt sich auch konzeptionell am Dienst bwDataArchiv. Es entwickelt eine Lösung zur Gewährleistung und effizienten kontinuierlichen Überprüfung der Datenintegrität des Archivs.

Als Speicherinfrastruktur nutzt bwDataArchiv ein HPSS-System (High Performance Storage Management) mit theoretisch unbegrenztem Speicherplatz. Das Datenmanagementsystem HPSS kommt unter anderem bei der Max Planck Computing and Data Facility (MPCDF), dem Deutschen Klima-

rechenzentrum (DKRZ) und verschiedenen führenden amerikanischen Rechenzentren zum Einsatz, welche auch die Weiterentwicklung des Systems verantworten. Die Infrastruktur ist hierarchisch aufgebaut und bietet, neben einem schnell zugreifbaren Plattenspeicher, in einer tieferen Speicherschicht Magnetbandspeicher, der am KIT über mehrere Bandbibliotheken an zwei Standorten mit Bandlaufwerken von unterschiedlichen Herstellern verteilt ist. Die HPSS-Architektur gewährleistet eine beliebige Skalierung und einen breiten Einsatzbereich, sowohl als Speicher mit hohem Durchsatz, als auch beim Dienst bwDataArchiv, als eine verlässliche und wirtschaftliche Datenablage.

bwDataArchiv kann als Speicherkomponente für ausgefeilte Langzeitarchivierungslösungen begriffen werden. Der Dienst bietet den Forschenden aber auch die Möglichkeit initiativ und eigenverantwortlich Daten auf lange Sicht zugreifbar zu machen. Gewährleistet wird dies durch ein sicheres Speicherverfahren, Bitstream Preservation genannt, das neben der Kontrolle der Daten durch Checksummen, eine sichere Übertragung und eine redundante Ablage auf Magnetbändern vorsieht. Im Vordergrund steht die transparente Einbindung der Archivierungskomponente in den Speicherungs- und Datensicherungsworkflow des Nutzers. Für bwDataArchiv kommen das universelle Übertragungsprotokoll sftp<sup>4</sup> und die im Grid-Umfeld verwendete Variante GridFTP zur Anwendung. Für die Übertragung mit GridFTP bietet bwDataArchiv mit der Eigenentwicklung bwDAHub eine vorkonfigurierte Lösung auf Basis von uberftp und globus-url-copy. Die grobe Architektur der bwDataArchiv-Infrastruktur zeigt Abb. 3.

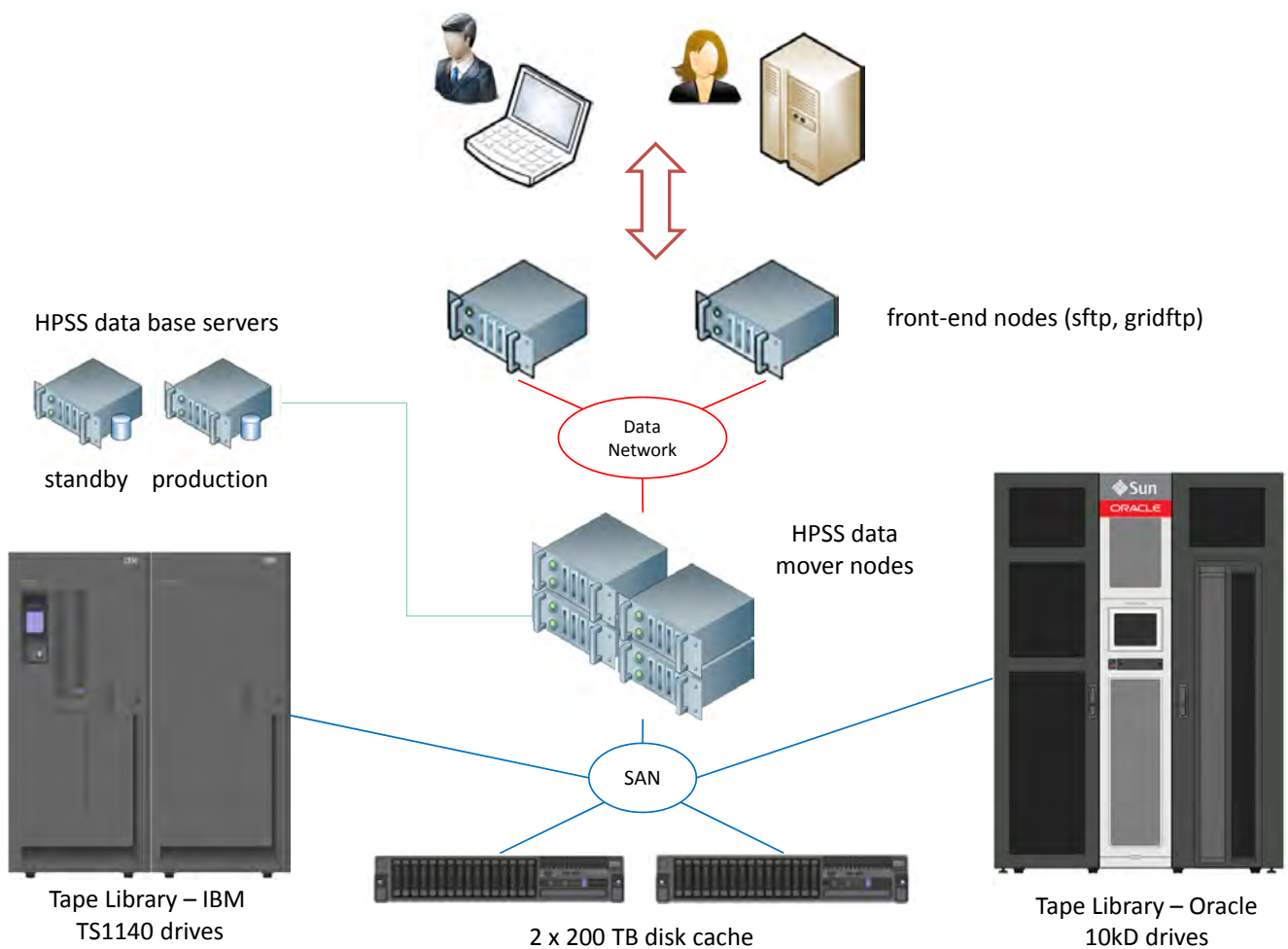


Abb. 3: Schematische Darstellung der bwDataArchiv-Infrastruktur

Eine geplante Weiterentwicklung von bwDataArchiv sieht eine einfache Publikationskomponente für automatisch erzeugte Daten aus Messinstrumenten und die Funktionalität des begrenzten Austauschs von Forschungsdaten in Arbeitsgruppen vor. Die Anforderungen für die Entwicklung dieser Komponenten basieren auf Nutzerwünschen, die unter anderem über das Landesprojekt bwFDM-Communités erhoben wurden. Das Landesprojekt bwDIM (bwDataInMotion) bei dem das SCC, die KIT-Bibliothek und das Leibniz Institut für Informationsinfrastruktur (FIZ) zusammenarbeiten, gewährleistet eine Weiterentwicklung des Dienstes und bindet die E-Science-Initiativen des

Ministeriums für Wissenschaft und Kunst (MWK) in das Langzeitspeicherprojekt mit ein. Der Zugang zum Dienst bwDataArchiv erfolgt über das föderierte Identitätsmanagementsystem bwIDM. Nutzer außerhalb der Föderation erhalten den Zugang z.B. über eine, von einem KIT-Mitarbeiter initiierte, persönliche Einladung via System-Mail.

Als Landesdienst wird bwDataArchiv ab 1.1.2017 Nutzern aus baden-württembergischen Hochschulen und deren Projektpartnern zur Verfügung stehen.

Weitere Informationen findet man unter [www.rda.kit.edu](http://www.rda.kit.edu). Beim Aufbau des

Dienstes haben mitgewirkt: Thomas Bönnisch, Frank Scheiner und Björn Schembera (HLRS). Felix Bach, Bettina Bauer, Ahmad Hammad, Dorin-Daniel Lobontu und Jos van Wezel (SCC).

### State-wide Service bwDataArchiv goes into Operation

With the bwDataArchive service research data can be securely stored for many years. Scientists are offered an easy to use platform that integrates with daily scientific workflows and offers a solution to their data management requirements. The service that builds on the HPSS storage system, is available for members of institutes in the state of Baden-Wuerttemberg and will be used as reliable bit storage by several projects and high level archival applications.

<sup>1</sup> Höchstleistungsrechenzentrum Stuttgart

<sup>2</sup> Research Data Repository (s. SCC-News 2015, S. 25)

<sup>3</sup> Archivierung von Dissertationen (s. SCC-News 2/2014, S. 37)

<sup>4</sup> secure file transfer protocol

## Virtuelle Graphikworkstations im Betrieb – ein Projektazit

Moderne Röntgentomographie, wie sie mit der ANKA-Synchrotronstrahlungsquelle durchgeführt wird, liefert enorme Datenmengen. Die zeitliche und räumliche Auflösung der Bilder erfordert leistungsfähige Analyseumgebungen, die nicht nur in der Lage sind, hohe Datenraten zu verarbeiten, sondern diese auch grafikbeschleunigt in 2D/3D zu analysieren. Im Rahmen des Projekts ASTOR wurde eine innovative Hard- und Softwareumgebung geschaffen, die Wissenschaftlern virtualisierte Graphikworkstations mit OpenGL-fähigem Remotezugang ad hoc zur Verfügung stellt. Diese einmalige Umgebung ist seit Anfang 2015 in Betrieb und wird sogar von Gastwissenschaftlern der Pennsylvania State University genutzt.

*Dr. Matthias Bonn, Dr. Wolfgang Mexner<sup>1</sup>*

### Virtuelle Graphikworkstations

Das Ziel des (inzwischen abgeschlossenen) ASTOR-Projektes liegt darin, wissenschaftliche Analysesoftware in virtualisierten Arbeitsplätzen, ähnlich einer echten Workstation, schnell und unkompliziert zur Verfügung zu stellen. Die Darstellung von 2D/3D-Daten, wie sie die Röntgentomographie erzeugt, erfordert leistungsfähige Graphikhardware, die in

den üblichen virtuellen Infrastrukturen nicht vorhanden ist, und Spezialsoftware, die in typischen virtuellen Serverinstanzen nicht läuft und auch nicht mit Webtechniken verfügbar gemacht werden kann.<sup>2</sup> VM-Ware-Spezialisten des SCC haben deswegen zusammen mit Wissenschaftlern der KIT-Institute IBPT und IPS einen VMware vSphere-Cluster aufgebaut, der hardwareseitig mit den notwendigen Graphikbeschleunigern ausgestattet ist<sup>3</sup>

und ein Portal [1] entwickelt, das die notwendige Orchestrierung des Clusters anstößt und alle Konfigurations- und Migrationsvorgänge der sehr komplexen VMware-API vor dem Nutzer versteckt.<sup>4</sup> Aus Nutzersicht ist lediglich eine von verschiedenen Vorlagen für virtuelle Computer zu selektieren (VM-Templates), mit der gewünschten CPU-, RAM- und Plattenkapazität zu konfigurieren und dann zu starten (Abb. 1). Neben speziellen ASTOR-Templates werden allgemein

gehaltene Templates, auch ohne Graphikbeschleunigung, in verschiedenen Linux- und Windowsvarianten angeboten, so dass auch weniger spezielle VM-Wünsche abgedeckt werden können. Hierbei können verschiedene, auch geographisch verteilte, vSphere-Installationen integriert werden, so dass Nutzergruppen voneinander separierte Umgebungen nutzen können. Der Instanzierungsprozess einer neuen VM im ASTOR-Cluster dauert, abhängig vom Template und der gewählten Festplattengröße, 5-15 Minuten. Danach

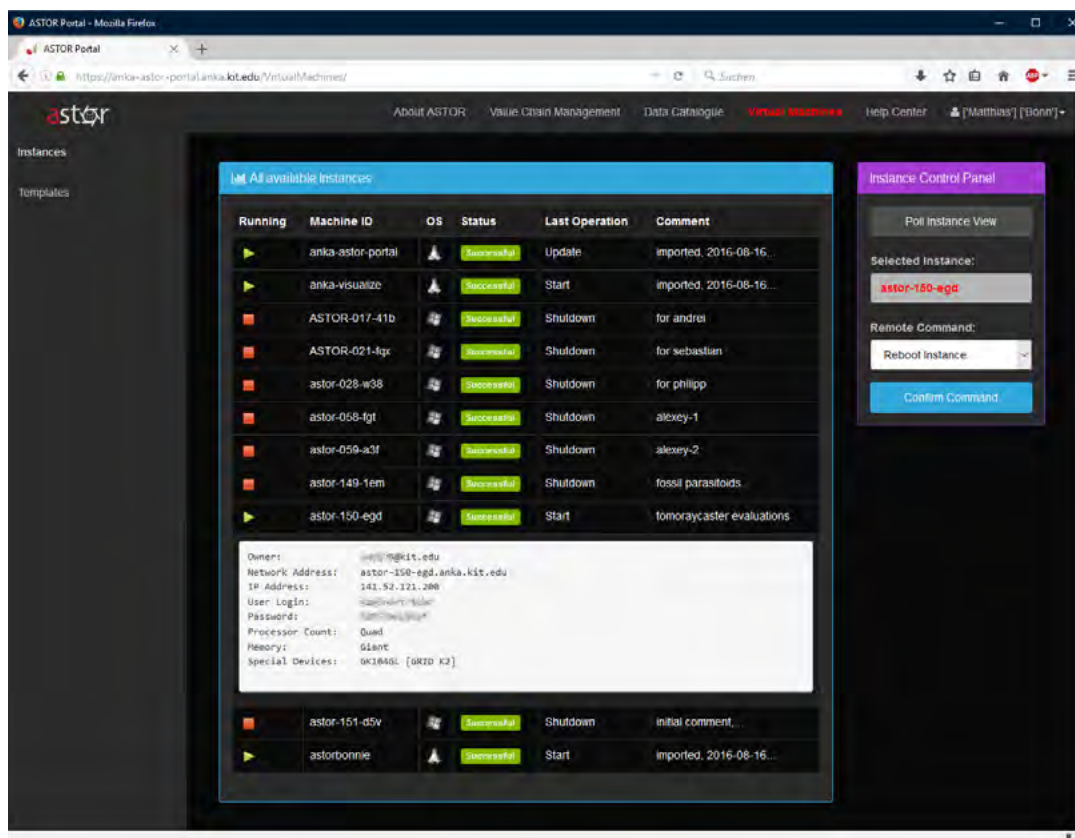


Abb. 1: Virtuelle Maschinen im ASTOR-Portal



ist der Login mit einem speziellen Client (VMware Horizon), der den grafikbeschleunigten Windows-Desktop per OpenGL oder DirectX remote zugänglich macht, möglich, und die vorinstallierte Analysesoftware kann genutzt werden. Der Systemzugang zu den Standard VMs ohne GPU geschieht mit den üblichen Bordmitteln Secure Shell (SSH) bzw. Remote Desktop Protocol (RDP).

Die für den Nutzer generierten Rohbilddaten der Röntgentomographieexperimente werden der VM automatisch über ein nutzerspezifisches Netzlaufwerk mit einer 10 GBit/s-Netzwerkanbindung eingeblendet. Entsprechende Mountpunkte und deren Einbindung sind in den Templates voreingestellt. So steht innerhalb weniger Minuten eine Arbeitsumgebung (wie in Abb. 2) zur Verfügung, für deren Nutzung lediglich ein Standard-Notebook, eine gewöhnliche 16 MBit DSL-Anbindung ans Internet zur Remote-Bildübertragung und ein KIT-Benutzerkonto notwendig sind. Der Registrierungsprozess läuft dabei zweistufig ab: Zunächst melden sich die Nutzer am Portal für Föderierte Dienste am KIT (bwidm.scc.kit.edu) für den ANKA-Dienst an, so dass die Bilddatenspeicher für den Nutzer ansprechbar werden. In einem zweiten Schritt werden die entsprechenden

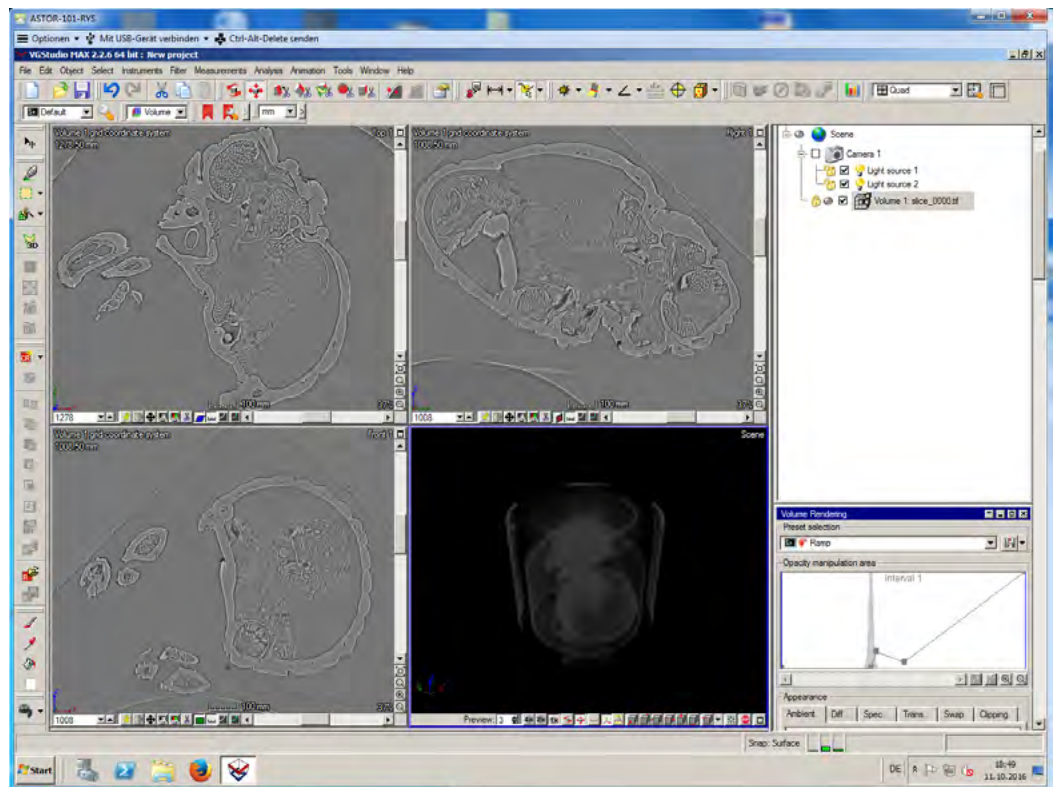


Abb. 2: Remotezugriff auf eine VM mit der Analysesoftware VGStudio MAX

Konten im VMware-Orchestrierungsdienst für den ASTOR-Cluster freigeschaltet. Die Authentifizierung der Nutzer an den Webportalen geschieht über Shibboleth Web-SSO. Die Authentifizierung an den OpenGL-Gastsystemen wird mit dem LDAP-Fassadenkonto durchgeführt, bei den generischen Gastsystemen mit lokalen root/Administratorrechten.

### Nutzungserfahrungen

Seit Anfang 2015 wurden die virtuellen Graphikworkstations von ca. 15 Nutzern – überwiegend Biologen – in Anspruch genommen. Unter anderem waren Gast-

wissenschaftler der Pennsylvania State University beteiligt, wobei zeitweise bis zu 25 ASTOR-VMs existierten. Da jedoch nur maximal acht davon gleichzeitig mit Graphikprozessor (GPU) betrieben werden können, wurde ein Mechanismus eingerichtet, der dafür sorgt, dass nicht benötigte VMs heruntergefahren werden, so dass die genutzten Ressourcen freigegeben und von anderen VMs genutzt werden können.<sup>6</sup> Dieses Verfahren ist ein Accounting und Billing mit sogenannten virtuellen Euros. Hierzu wird die Fähigkeit des VM-Orchestrierungsdienstes genutzt, im einstelligen Minutentakt bis zu ca. 25 Ressourcenverbrauchsmetriken pro VM zu

<sup>1</sup> Institut für Beschleunigerphysik und Technologie (IBPT) am KIT

<sup>2</sup> Wie z. B. die im Projektkontext eingesetzten Amira oder VG Studio MAX auf Basis von Windows Server.

<sup>3</sup> Die Projektumgebung besteht aus zwei ESXi Hosts, die zusammen 16 3 GHz Xeon Kerne, 700 GB RAM und 8 NVIDIA GRID K2 GPU Kerne anbieten. Für die VM-Images stehen 21 TB NFS-Speicher zur Verfügung.

<sup>4</sup> Technisch kommt PCI-Passthrough zum Einsatz, das ein physisches PCI-Device direkt in eine VM einblendet, weswegen verschiedene VMware-Automatismen wie DRS und vMotion für die betreffenden VMs nicht nutzbar sind und die Zuordnung von VM zu Host und PCI-Device selbst durchgeführt werden muss.

<sup>5</sup> Das Mapping erfolgt hier über den Shibboleth-internen Identifier ePPN (eduPersonPrincipalName)

<sup>6</sup> Die bei VMware normalerweise problemlose Überprovisionierung von Laufzeitressourcen, wie RAM, steht bei laufenden VMs mit PCI-Passthrough-Hardware nicht zur Verfügung. Eine gestartete VM erhält ihre Ressourcen immer exklusiv reserviert. Stehen die Ressourcen nicht zur Verfügung, kann die VM nicht gestartet werden.

erfassen und über längere Zeiträume, u. a. in Stunden- und Tagesaggregationen, zu archivieren. Darauf aufbauend ist ein Kostenmodell<sup>7</sup> erstellt worden, das den Ressourcenverbrauch auf Stundenbasis in virtuelle Euros umrechnet und bei Überschreiten eines monatlichen Grundbudgets zunächst den betreffenden Nutzer informiert und später dann die VMs des Nutzers mit steigender Frequenz abschaltet. Das definierte Kostenmodell erlaubt es, eine typische GPU-VM zehn Stunden pro Tag zu nutzen, dabei bleibt es dem Nutzer überlassen, wie er das jeweilige Kontingent einsetzt. Ein Kosten-simulator erlaubt es vorab, die entstehenden „virtuellen Kosten“ abzuschätzen. Die Billing- und ggf. Shutdown-Prozesse

werden vollautomatisch durchgeführt, eine einmalige oder dauerhafte Erhöhung des monatlichen Budgets ist jedoch zusätzlich möglich. Dieser Mechanismus erhöhte die Disziplin der Nutzer. Durch das veränderte Verhalten und den Zeitzoneversatz der amerikanischen Anwender, konnten deutlich mehr Forscher bedient werden, als es mit den verfügbaren Ressourcen oder gar physischen Workstations machbar gewesen wäre.

### Fazit

Die Implementierung der ASTOR-Systemumgebung stellt einzigartige Funktionen zur Verfügung und bietet, im Vergleich zu traditionellen IT-Umgebungen mit lokalen

Workstations, signifikante Verbesserungen. Wissenschaftler können somit ein flexibles System von überall aus nutzen, das zudem wesentlich leichter auf einem aktuellen technischen Stand (Hard- und Software, inklusive Lizenzierung) gehalten werden kann. Dies gilt nicht nur für die spezialisierte ASTOR-Umgebung, sondern auch für damit bereitgestellte VMs, die dem generellen Einsatz in Forschung und Lehre dienen. Darüber hinaus könnten die VMs feingranular monetär abgerechnet werden.

### Weitere Informationen

[1] <https://astor.kit.edu>

<sup>7</sup> Beispiel für die flexible virtuelle Abrechnung: 0.007 € pro GB RAM + 0.05 € pro CPU Core + 0.5 € pro nVidia GPU Core, stündlich und nur bei laufender VM. VM-Imagespeicher und die genutzte CPU-Takte werden erfasst, aber nicht eingerechnet.

### Working with Virtual Graphics Workstations – Results of the ASTOR-Project

Modern X-ray tomography, as provided by the ANKA synchrotron light facility, results in immense data amounts, the temporal and spatial resolution of the resulting 3D images demands for powerful environments. These hard- and software must be able to process the high data rates and needs the capability to analyze the image data with help of accelerated graphics processing in 2D/3D. Within the ASTOR project, the SCC has implemented together with scientists of IBPT and IPS a system providing scientists virtualized workstations with an OpenGL accelerated graphics subsystem, remotely usable from everywhere in the Internet. The environment is in use since 2015, is usable cloud-like as-a-service, and it is currently also evaluated by guest scientists from Pennsylvania State University.

## ILIAS 5 – Update des E-Learning-Systems des KIT

Die neue Version 5 des am SCC betriebenen E-Learning Systems ILIAS wurde im September 2016 eingespielt. Mit ILIAS 5 wurden nicht nur notwendige Sicherheits-Updates durchgeführt, es stehen nun auch neue Funktionen zur Verfügung.

*Franziska Wandelmaier, Alvar Wenzel*

### Neuerungen

ILIAS 5 verfügt über eine erweiterte Funktionalität<sup>1</sup>, die z.T. auch speziell von den ILIAS-Nutzerinnen und Nutzern am KIT gewünscht wurde, z.B.

- Automatisches Nachrücken von Wartelisten
- Laden auch mehrerer Dateien gleichzeitig in Kursräume per Drag & Drop
- schnelles Aufräumen des „Persönlichen Schreibtischs“ durch gleichzeitiges Löschen mehrerer Objekte
- erweiterte Sortiermöglichkeiten, etwa innerhalb von Gruppen und Kategorien
- PDF-Export von Wikis (für bestimmte Seiten oder insgesamt)
- erweiterte Link-Möglichkeiten in Wikis
- Zuordnung beliebig vieler Kursbetreuer mit Kontakt-Möglichkeit über die Info-Seite des Kurses
- gezielter Versand von E-Mails aus Kursen, etwa nur an Administratoren oder nur an Tutoren

Es ist geplant über weitere Highlights der neuen ILIAS-Version auf der ILIAS-Startseite als „Tipp des Tages“ zu informieren. Im Zuge der Weiterentwicklung von ILIAS wurde mit der Version 5 auch das äußere Erscheinungsbild moderner gestaltet, beispielsweise durch neue Icons (s. Abb. 1). Außerdem wurde das Layout für mobile Anzeigeräte weiter optimiert.

### Was bedeutet die Umstellung für die ILIAS-Nutzerinnen und Nutzer?

Grundsätzlich ändert sich für die Nutzerinnen und Nutzer von ILIAS durch die Umstellung wenig. Es hat sich gezeigt, dass die meisten Nutzer sich nach kurzer Zeit an das neue Erscheinungsbild gewöhnen konnten (Vergleich s. Abb. 1) und keine gravierenden Probleme bei der Bedienung auftraten.<sup>2</sup>

Alle bereits in ILIAS 4.3 vorhandenen Kurse wurden automatisch in die neue Version 5.1 überführt. Die Dozierenden selbst muss-

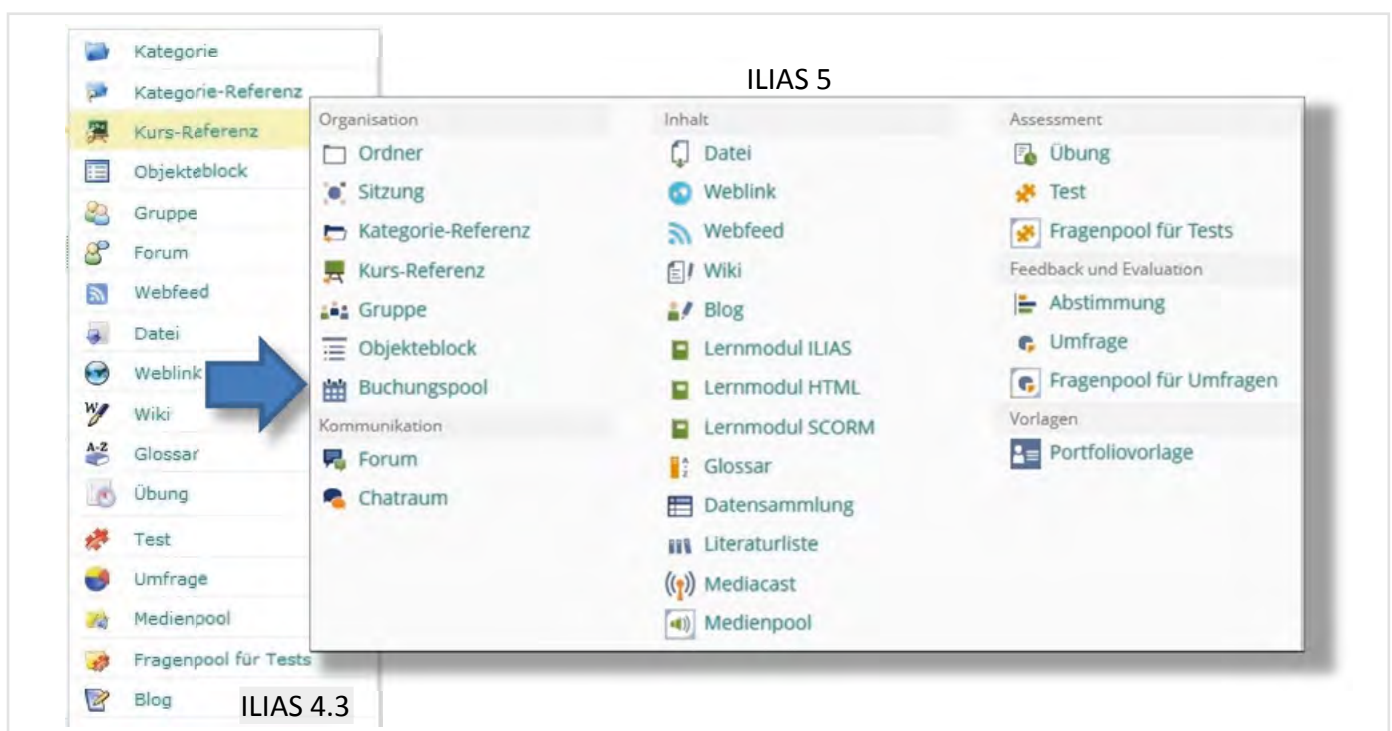


Abb. 1: Neue Funktionen im veränderten Layout von ILIAS 5 (rechts); daneben zum Vergleich die Ansicht in ILIAS 4.3

<sup>1</sup> Zu den Neuerungen siehe auch das Feature Wiki von ILIAS: [http://www.ilias.de/docu/goto.php?target=wiki\\_1357\\_Overview&lang=de](http://www.ilias.de/docu/goto.php?target=wiki_1357_Overview&lang=de)

<sup>2</sup> Um den ILIAS-Nutzerinnen und Nutzern den Umstieg zu erleichtern, bietet das Zentrum für Mediales Lernen (ZML) Schulungen, Webinare und weitere Informationsmaterialien zur neuen Version 5 an: <http://www.zml.kit.edu/e-learning-schulungen.php>.

ten daher keine Kurse übertragen. Bei dem Update wurde – bildlich gesprochen – der Antrieb gewechselt und mit mehr Funktionalität ausgestattet; das damit betriebene System bleibt jedoch in der Form erhalten, in der es Dozierenden und Studierenden aus der Vergangenheit bekannt ist.

### Ablauf der Umstellung

Beim Update auf ILIAS 5 wurden über 14.000 Kurse und ca. 65.000 Benutzerkonten transferiert. Das Installationsverzeichnis auf dem Filesystem (ca. 1 Terabyte) musste für das Update auf eine andere Versionsverwaltung (von SVN auf Github) umgestellt werden. Die komplette Datenbank (mit einer Größe von über 30 Gigabyte) wurde umstrukturiert. Der Update-Prozess verlief zwar weitgehend automatisch, war aber zeitaufwändig und erforderte an vielen Stellen manuelles Eingreifen, da sich über die lange Laufzeit der bestehenden ILIAS-Installation an mehreren Stellen Unstimmigkeiten und sogar Fehler in der Datenbank eingeschlichen hatten.

Aufgrund des großen Versionssprungs, von ILIAS 4.3 auf ILIAS 5.1, mussten während des Update-Prozesses einige Datenbanktabellen von Hand hinzugefügt, gelöscht oder nachbearbeitet werden. Diese Arbeiten erforderten einen langen Vorlauf mit vielen Vorarbeiten, zahlreichen kleineren Test-Update-Sequenzen und einem großen Test-Update an einer Spiegelung des Produktsystems.

Die Anpassungen des Layouts für das KIT waren in der neuen ILIAS-Version besonders zeitaufwändig, da sich an dieser Stelle die zur Verfügung stehenden Konfigurations-Möglichkeiten und der gesamte Anpassungsprozess wesentlich verändert haben.

### ILIAS 5 – Update of KIT's E-Learning System

KIT's e-learning-system ILIAS, which is hosted at SCC, was updated to version 5 in September 2016. ILIAS 5 not only includes important security updates, it also offers several new features (such as drag-and-drop for uploading multiple files) and is compatible with a larger set of plugins. The layout has also changed in version 5 with modernized icons and design. Internal processes were optimized, including database queries and caching mechanisms, to speed up ILIAS.

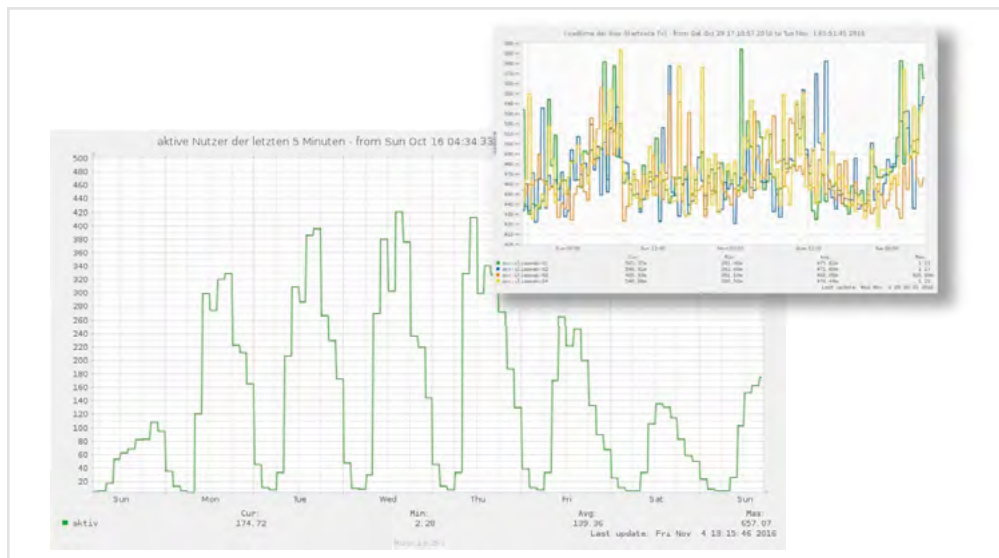


Abb. 2: Ladezeit der ILIAS-Startseite im Wochenüberblick (rechts oben, Maximum: 1,21 Sekunden, Mittel: 0,47 Sekunden) bei stark wechselnder Zahl gleichzeitig aktiver Nutzer (Maximum: 657 Nutzer, Mittel: 139)<sup>4</sup>

Vor der Freigabe der neuen Version für alle Nutzer wurden intern zahlreiche, im Vorfeld erstellte Testszenarien durchgespielt.

### Weitere Optimierungen

In der neuen ILIAS-Version stehen verschiedene Caching-Methoden (wie z.B. Memcached) zur Verfügung. Durch deren Einsatz, kombiniert mit einer verbesserten Programmierung des ILIAS-Source Codes (beispielsweise bei Datenbankabfragen) und dank der guten Aufstellung unserer mittlerweile komplett virtualisierten ILIAS-Installation<sup>3</sup>, können nun auch plötzliche Nutzeranstürme vom System besser bewältigt werden (s. Abbildung 2).

Diese Verbesserung macht sich besonders zu Vorlesungsbeginn bemerkbar, da Sprünge von so gut wie keinem angemeldeten Nutzer auf knapp 600 gleichzeitig aktiven Nutzern innerhalb weniger Minuten keine Seltenheit darstellen.

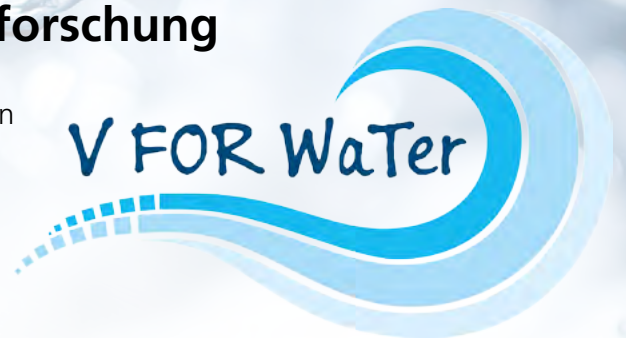
Des Weiteren eröffnen sich mit der neuen ILIAS-Version Möglichkeiten zur Anbindung aktuell unterstützter Plugins an die ILIAS-Installation des KIT, wie z.B. LiveVoting und der bald schon zur Verfügung stehenden Anbindung an die Bibliothek des KIT.

Fragen zum neuen ILIAS 5 beantwortet gerne der ILIAS-Support ([ilias@studium.kit.edu](mailto:ilias@studium.kit.edu)).

<sup>3</sup> Vier Webserver, eine Service-Maschine sowie ein virtueller Datenbankserver (s. auch SCC News 2015, S. 12-13).

<sup>4</sup> Die Grafiken zeigen bereits auf bestimmte Zeitintervalle gemittelte Werte, sodass die genannten Maxima in der Grafik nicht mehr zu erkennen sind.

# V-FOR-WaTer – Eine virtuelle Forschungsumgebung für die Wasser- und terrestrische Umweltforschung



Das Management sowie die Aufbereitung von heterogenen Datenbeständen für hydrologische Analysen und Modellierungen stellt oftmals eine große Herausforderung dar. Typischerweise erfordert dies ein zeitintensives Umformatieren unterschiedlichster Daten verschiedenster Quellen für die jeweilige Fragestellung. Die virtuelle Forschungsumgebung V-FOR-WaTer ermöglicht einen einfachen Zugriff auf die Daten im benötigten Format und bietet ebenso eine Plattform, auf der Analysewerkzeuge gespeichert, geteilt und genutzt werden können. Um den Anforderungen der Umweltforschung gerecht zu werden, erfolgt die Entwicklung der Forschungsumgebung in Zusammenarbeit mit dem Institut für Wasser und Gewässerentwicklung (IWG) des KIT.

*Dr. Marcus Strobl, Dr. Sibylle Hassler<sup>1</sup>*

„Startschuss“ des zunächst für zwei Jahre finanzierten Landesprojektes V-FOR-WaTer war am 1.7.2016, wobei bereits im Februar Abstimmungsgespräche mit den Universitäten in Baden-Württemberg und der Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW) stattfanden. Durch die frühe und intensive Zusammenarbeit mit der wissenschaftlichen Gemeinschaft in diesem Forschungsbereich wird die bestmögliche Akzeptanz und Nutzbarkeit der Virtuellen Forschungsumgebung (VFU) erreicht.

Der Schwerpunkt von V-FOR-WaTer liegt auf der Haltung hydro-meteorologischer Daten und dem direkten Zugriff auf Analysewerkzeuge, um Forschungsdaten schnell prozessieren zu können und für die prognostische Modellierung nutzbar zu machen. Die erste Phase ist auf die Integration von Wasser- und Umweltdaten aus Baden-Württemberg fokussiert.

Insbesondere der Oberrheingraben ist durch seine lange Geschichte in der Umweltbeobachtung und der großen Diversität der Landschaftsräume und Landnutzungen prädestiniert für den Aufbau eines Langzeitobservatoriums mit einer gemeinsamen VFU. Die Kernidee besteht

im zielgerichteten und abgestimmten Ausbau bestehender Forschungseinzugsgebiete der Universitäten unter der Integration des Umweltmessnetzes der Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW) sowie der Anbindung des Deutschen Wetterdienstes (DWD). In diesem Observatorium betreiben die Universitäten Baden-Württembergs und die LUBW, wissenschaftlich koordiniert, ihre Umweltbeobachtungsmessnetze. Die gemessenen Daten können so gemeinsam genutzt werden, um insbesondere langfristige Auswirkungen des globalen Wandels besser zu verstehen. Dies ist ein wichtiger Schritt hin zur Realisierung der Vision eines „Landesobservatoriums Oberrheingraben“.

Die Speicherung der Modellparametrisierungen und Simulationsergebnisse in V-FOR-WaTer für unterschiedliche Gebiete, und die Entwicklung geeigneter Dokumentationsstandards, ermöglicht es die Erkenntnisse aus den Modellen zu bewahren und zu teilen. Somit kann langfristig ein vergleichendes Lernen über die Vor- und Nachteile unterschiedlicher Modellsysteme für die Anwendung in unterschiedlichen Regionen und für unterschiedliche Ziele gewährleistet werden.

Langfristig soll V-FOR-WaTer auch über Landesgrenzen hinaus als gutes Beispiel für den Umgang mit Umweltdaten gelten und erweitert werden. Dies wird durch eine modulare Ankopplung an andere, bereits bestehende Systeme, wie z.B. das der LUBW, aber potentiell auch nationale oder internationale Systeme, gewährleistet.

## Weitere Informationen

<http://www.vforwater.de>

### V-FOR-WaTer – a virtual research environment for water and terrestrial environmental research

Management and preprocessing of heterogeneous data pools for hydrological analyses and models often bears a big challenge. Typically, a time-consuming transformation of different data from various sources is required. V-FOR-WaTer provides simple access to data in the required format and provides a platform to save, share and use tools for data analysis. To meet the various requirements of environmental research, the V-FOR-WaTer virtual research environment is developed in close cooperation with the Institute of Water and River Basin Management at KIT.

<sup>1</sup> Institut für Wasser und Gewässerentwicklung (IWG)

# EUDAT Operations Management

Das EUDAT Konsortium ist eine internationale Kooperation, die digitale Dateninfrastrukturen, Ressourcen und Datendienste fördert, um Forschungs- und Wissenschafts-Communities beim Datenmanagement zu unterstützen. 35 europäische Organisationen nutzen Dienste und Speicherressourcen in 15 Ländern, die eine sogenannte kollaborative Dateninfrastruktur (CDI) bilden. Ziel ist es, wissenschaftliche Daten langfristig zu speichern, aufzufinden und in einer vertrauenswürdigen Umgebung zu verarbeiten. Durch die Einführung eines Operationsteams sowie geeigneter Managementprozesse wird die Kommunikation zwischen den Partnern auf technischer Ebene wirkungsvoll gebündelt. Das Operations Management bildet eine wichtige Schnittstelle in der CDI, die aus 13 generischen Datenzentren und einer wachsenden Anzahl von Wissenschaftsgemeinschaften und themenspezifischen Zentren besteht. Im Rahmen des Arbeitspaketes "Operations" koordiniert das SCC die Implementierung und den Betrieb von Diensten, sowie die Bereitstellung der Ressourcen.

*Pavel Weber*

## EUDAT – Datendienste für die Wissenschaft

EUDAT2020, das dreijährige Nachfolgeprojekt von EUDAT, umfasst sieben Kern-Communities - CLARIN, ELIXIR, ENES, EPOS, ICOS, LTER und VPH. Diese repräsentieren ein sehr breites Spektrum aus Sozial- und Geisteswissenschaft, Geographie, Energie und Umweltwissenschaft, aber auch Biologie und Biowissenschaft. Für diese sehr heterogenen Wissenschafts-Communities bietet EUDAT IT-Dienste, die es ermöglichen wissenschaftliche Daten effizient zu verwalten sowie hochverfügbar und langfristig zu speichern. Alle EUDAT-Aktivitäten zur Erbringung dieser IT-Dienste sind in Arbeitspaketen organisiert. Hier wird das Arbeitspaket Operations näher betrachtet.

## Inbetriebnahme der CDI

Innerhalb des Arbeitspaketes "Operations" werden Konzepte entwickelt, um die CDI technisch zu konsolidieren. Ziel ist es die Integrität, die Vertraulichkeit und die sichere Aufbewahrung der von den Benutzern hinterlegten Daten zu gewährleisten. Hierfür wurde eine neue, verteilte IT-Infrastruktur aufgebaut. Diese besteht aus 13 generischen Datenzentren und einer wachsenden Anzahl an Wissenschafts-Communities sowie thematischen Zentren. Im letzten Jahr konzentrierten sich die Aktivitäten von "Operations" auf die Umsetzung der Datenprojekte auf die CDI. Diese Aufgabe erfordert die Zusammenarbeit aller Partner, die die Nutzer mit Werkzeugen für den IT-Betrieb, einer effektiven Kommunikation, durch Software-Monitoring und einen Helpdesk

unterstützt. Das Operationsteam soll garantieren, dass die erforderlichen Ressourcen rechtzeitig zur Verfügung gestellt werden können.

## Blick in die Projektstruktur

Das breite Spektrum von notwendigen Maßnahmen wurde in fünf Aufgabenbereiche zusammengefasst (Abb. 1)

**Data Project Consultancy & Enabling:** Hierunter fällt der Aufbau von Expertenwissen zur Unterstützung der Forscher und Anwender bei der Nutzung der CDI für deren Datenprojekte.

**Helpdesk & Support:** Diese Task unterstützt die EUDAT CDI Benutzer durch ein zentrales Helpdesk-System und definiert die dafür notwendigen Richtlinien. Die einlaufenden Anfragen verteilt ein „Dispatching Officer“ über das Supportsystem an die Partnerorganisationen, die den 2nd-Level-Support leisten.

**Security:** Dieses Aufgabenpaket konzentriert sich auf die Koordinierung der operativen Sicherheit und auf die Sicherheitsbewertung, die Vertraulichkeit, Integrität und Zugänglichkeit der Daten, der Dienste und der Systeme von EUDAT.

**Operational Tools & Services:** Bereitstellung zentraler Basisdienste wie z.B.

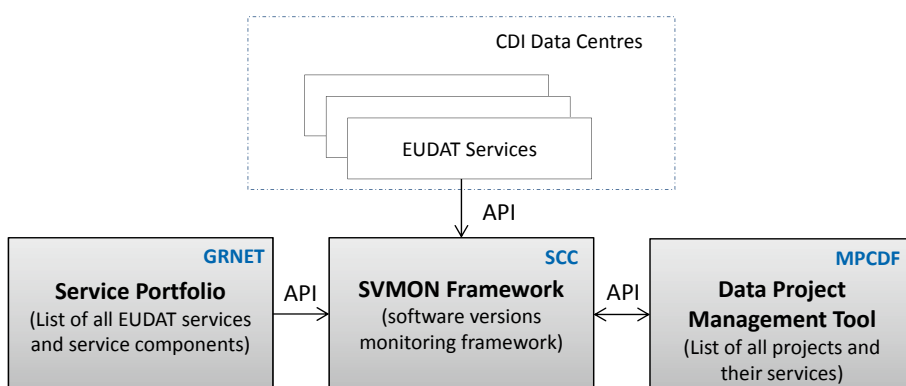


Abb. 2: Schema der automatisierten Software-Versionen-Überwachung innerhalb der CDI.

Accounting, Überwachung, Inventarisierung und Authentifizierung, die benötigt werden, um die CDI zu betreiben.

Service and Resource Provisioning & Coordination: Über diese Tasks soll die von den Datenprojekten erforderliche Überwachung der Produktivität, der Servicequalität, der Sicherheit und Nachhaltigkeit der CDI koordiniert werden. Eine weitere Aufgabe ist die Bereitstellung der IT-Ressourcen und zugehörigen Services.

**Clevere Betriebsprozesse**

Das SCC verantwortet den Bereich Service and Resource Provisioning & Coordination. Die vom Operations Manager des SCC koordinierten übergreifenden Verfahren für die Planung, Steuerung und den Betrieb der Infrastruktur und Dienste basieren auf den Service Management Standards FitSM ([www.fitsm.eu](http://www.fitsm.eu)).

In Rahmen des Arbeitspakets Operations wurden aus den 14 FitSM-Kernprozessen die für den Service-Support wichtigen Prozesse – Incident & service request, Problem, Configuration, Change und Release & deployment management – für EUDAT angewendet, sowie dafür passende Richtlinien definiert. Für die erfolgreiche Koordination des EUDAT Betriebs ist eine tägliche Interaktion mit den anderen oben aufgeführten Aufgabenbereichen erforderlich. Dies wird u.a. auch durch die Einführung cleverer und schlanker IT-Betriebsprozesse, wie sie FitSM definiert, erreicht.

**Werkzeuge und Workflows im Betrieb**

Eine so weitverteilte Infrastruktur, wie die

CDI, ist nur über geeignete Werkzeuge und Workflows effizient zu verwalten. Es ist wichtig durch deren Nutzung die Kommunikation per E-Mail oder Telefon zu reduzieren und stattdessen operative Daten automatisch zwischen den unterschiedlichen Betriebsprozessen zu übertragen. Des Weiteren soll die Vielfalt der Werkzeuge konsolidiert werden und alle Informationen mit Hilfe von APIs in einem einheitlichen Datenpool zusammenlaufen. Diesen Ansatz nutzt beispielsweise das "Software Version Monitoring" (SVMON), das im SCC entwickelt wurde. Es stellt ein Framework zur Verfügung, das regelmäßig die aktuellen Software Versionen aller EUDAT-Dienste abfragt und abspeichert und mit der Service Portfolio Datenbank vergleicht, ob Software-Updates notwendig sind (siehe Abb. 2). Die Ergebnisse werden an ein weiteres EUDAT-Werkzeug, das Data Project Management Tool, geschickt. Dieser Ansatz erlaubt es den unterschiedlichen Datenprojekten automatisch und schnell

veraltete Software zu identifizieren und effizienter zu aktualisieren.

**Ausblick**

Experten des SCC bieten neben der Koordination bei der Bereitstellung von IT-Diensten und Ressourcen auch technische Unterstützung zur Installation und Konfiguration von EUDAT-Diensten an. Für den EUDAT-Dienst B2SAFE wurde am SCC eine automatisierte Installation und Konfiguration, basierend auf dem Systemkonfigurationswerkzeug Puppet, entwickelt. Die Pakete für die automatische Installation weiterer Dienste wie z. B. B2STAGE werden derzeit geschrieben und stehen bald allen Datenzentren zur Verfügung.

**Weitere Informationen:**  
 SCC erweitert Beteiligung in EUDAT;  
 SCC-News 2015, S. 28

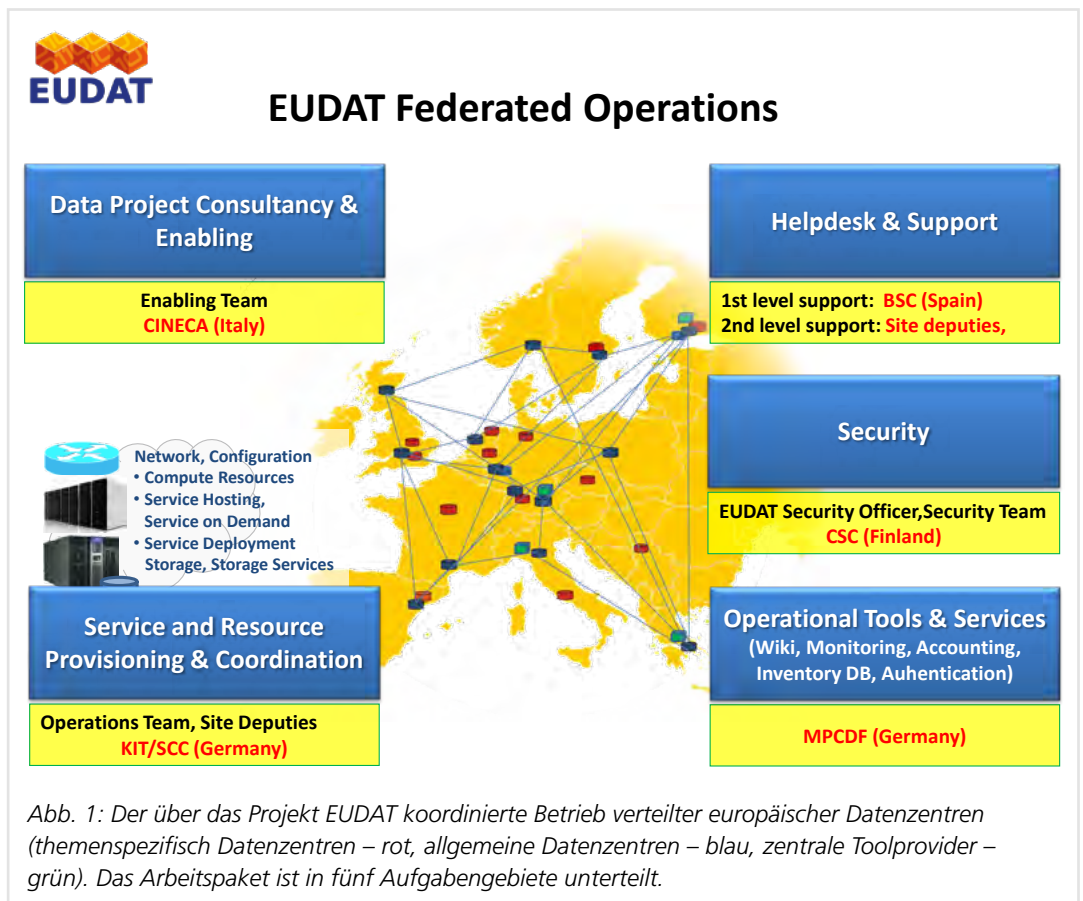


Abb. 1: Der über das Projekt EUDAT koordinierte Betrieb verteilter europäischer Datenzentren (themenspezifisch Datenzentren – rot, allgemeine Datenzentren – blau, zentrale Toolprovider – grün). Das Arbeitspaket ist in fünf Aufgabengebiete unterteilt.

**>>> EUDAT Operations Management**

The EUDAT consortium is an international collaboration, which federates the digital capabilities, resources, and data services to engage and support multiple research communities in efficient data management. A geographically distributed, resilient network of 35 European organizations shares service and storage resources across 15 countries and builds a Collaborative Data Infrastructure (CDI) to preserve, find, access and process data in a trusted environment. The EUDAT operations team organizes the co-operation of the different actors within the CDI and provides the distributed operational infrastructure with 13 generic data centers in partnership with a growing number of collaborating research communities, thematic centers and research infrastructures. The contribution of SCC to the EUDAT Operations task is focused on the coordination of deployment and operation of services as well as resource provisioning.

## Gründung der EUDAT Collaborative Data Infrastructure

Fünfzehn Forschungseinrichtungen haben die EUDAT Collaborative Data Infrastructure gegründet. Ziel ist der Aufbau einer zuverlässigen Dateninfrastruktur für das Management von Forschungsdaten in Europa.

*Prof. Dr. Achim Streit, Daniel Mallmann<sup>1</sup>, Stefan Heinzel<sup>2</sup>*

Fünfzehn Einrichtungen, darunter das Jülich Supercomputing Center (JSC), das Steinbuch Centre for Computing (SCC) und die Max Planck Computing & Data Facility (MPCDF), haben kürzlich die EUDAT Collaborative Data Infrastructure (EUDAT CDI) gegründet. Ziel der EUDAT CDI ist es, eine nachhaltige Perspektive für das Management von Forschungsdaten in Europa zu bieten und dafür eine zuverlässige Dateninfrastruktur zu etablieren und langfristig zu betreiben. Die Kollaboration besteht aus europäischen Datenzentren und Betreibern von Forschungsdatenrepositorien. Grundlage der Organisation ist eine partnerschaftliche Vereinbarung, die sich aus dem in Horizon 2020 geförderten EUDAT2020 Projekt entwickelt hat. Die EUDAT CDI bildet somit auch einen Grundstein zur Etablierung der European Open Science Cloud (EOSC). Im Rahmen des ersten EUDAT CDI Council Meeting, am 27.09.2016 in Krakau, wurde die EUDAT CDI offiziell ins Leben gerufen. Zeitgleich wurden drei neue Partner aufgenommen und weitere Organisationen haben ihr Interesse zum Beitritt erklärt.

Die Daten-Infrastruktur umfasst neben generischen und multi-disziplinären wissenschaftlichen Datenzentren, wie dem JSC und dem SCC, auch fachspezifische Zentren, beispielsweise das Deutsche Klimarechenzentrum. Die daten-orientierten Dienste aus den Bereichen „Data Sharing“, „Data Access“, „Data Storage“, „Data Preservation“, „Data Processing“ und „Data Discovery“ werden von Wissenschaftlern unterschiedlicher Disziplinen gemeinsam mit den Betreibern dieser Dienste entworfen und betreut. Auf große multi-disziplinäre Forschungseinrichtungen zugeschnittene Dienste sind genauso Bestandteil des Portfolios, wie solche, die sich an kleinere Forschungsgruppen und einzelne Wissenschaftlerinnen und Wissenschaftler wenden. Das JSC betreibt derzeit u.a. die zentrale Nutzeridentifikation und allgemeine Dienste zur Synchronisation und zum Austausch von Daten, das SCC und die MPCDF betreiben Dienste zur Langzeitspeicherung und Archivierung. Alle Zentren der EUDAT CDI unterstützen den Aufbau von Forschungsrepositorien gemäß den Grundsätzen von Open Access.

JSC is one of fifteen European partners that recently founded the EUDAT Collaborative Data Infrastructure (EUDAT CDI). The partners consist of European data centres and operators of community repositories. EUDAT CDI is a partnership agreement that evolved from the EUDAT2020 project, funded by the Horizon 2020 programme. The goal of EUDAT CDI is to offer a long-term perspective beyond current project funding for managing scientific data, thus becoming a reliable infrastructure for scientific communities and researchers. EUDAT CDI was launched at the first EUDAT CDI Council meeting, which took place on 27 September 2016 in Krakow. At this meeting, three new partner applications for joining EUDAT CDI were accepted.

The infrastructure comprises generic or multi-domain data centres, such as JSC, and community-specific data centres. The infrastructure services are co-designed by community scientists and data service developers. EUDAT CDI offers tailored services for large research communities and generic services for small research groups and individual scientists. JSC currently operates the central user identification, a generic service for data synchronization and exchange, and community-specific services.

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

<sup>2</sup> Max Planck Computing and Data Facility (MPCDF)

### Weitere Informationen:

<https://www.eudat.eu/eudat-cdi>



## bwDIM – Vereinfachter Datenfluss im Forschungsdatenmanagement



Das baden-württembergische Landesprojekt bwDataInMotion (bwDIM) unterstützt Wissenschaftlerinnen und Wissenschaftler der Universitäten in Baden-Württemberg beim Forschungsdatenmanagement. Es hat zum Ziel, den Datenfluss zwischen den unterschiedlichen Systemen zu vereinfachen.

*Felix Bach*

Verfügbarkeit und Nutzbarkeit wissenschaftlicher Daten sind Schlüsselfaktoren für die weitere Entwicklung der empirischen Wissenschaften. Das Management, der langfristige Erhalt und die Zugänglichmachung dieser Daten ist jedoch eine große Herausforderung für Hochschulen und außeruniversitäre Forschungseinrichtungen.

Das baden-württembergische Landesprojekt bwDIM (bwDataInMotion) ist eine Kooperation des Steinbuch Centre for Computing mit dem Leibniz-Institut für Informationsinfrastruktur (FIZ Karlsruhe) und der KIT-Bibliothek und zielt auf den vereinfachten Datenfluss zwischen den unterschiedlichen Systemen zum Forschungsdatenmanagement. Es soll damit den Arbeitsalltag für die Forschenden erleichtern, indem technische und organisatorische Hürden bei der Archivierung, Publikation und Nachnutzung von Forschungsdaten überwunden werden (s. Abb. 1).

Dabei wird auf bestehenden Bausteinen und Vorarbeiten aus anderen Projekten (u.a. bwDataArchiv, RADAR, bwDataDiss) aufgebaut. Diese werden zu einem größeren Ganzen verknüpft. Hierzu untersucht das Projekt die zentralen Arbeitsprozesse der Forscherinnen und Forscher und identifiziert bestehende Defizite auf Prozessebene und bei den technischen Basisdiensten. Darauf aufbauend werden Lösungsansätze zur Prozessoptimierung sowie zur Optimierung und Integration der Basisdienste erarbeitet, umgesetzt und anhand bestehender Systeme und Nutzergruppen validiert.

Konkret beinhaltet dies die Vereinheitlichung von Archivschnittstellen, Standardisierung administrativer und technischer Metadaten, föderationsübergreifende Authentifizierung und Autorisierung, Unterstützung multipler Datenübertragungs- und asynchroner Datenbereitstellungsverfahren sowie die Anbindung der technischen Infrastruktur der Verlags- und Repositoriensysteme.

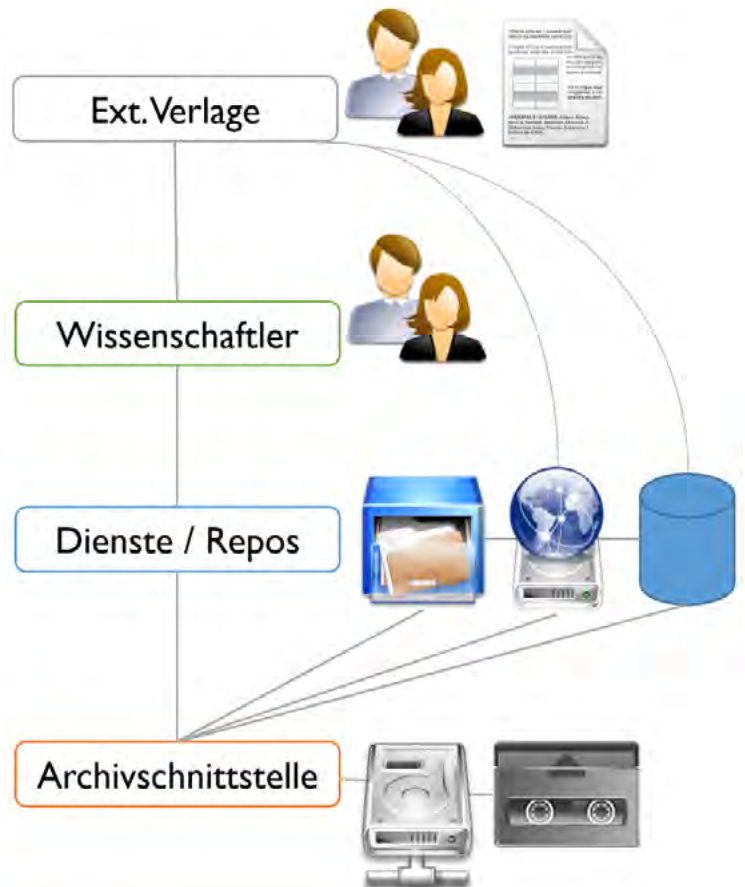


Abb. 1: Schematische Darstellung der Datenflüsse beim Forschungsdatenmanagement

### bwDIM – simplifying data flows in research data management

The project bwDataInMotion (bwDIM) supports scientists of universities in Baden-Württemberg in managing their research data. The goal of the project is to simplify the flow of data between the diverse storage- and repository systems. This includes the unification of archive interfaces, standardisation of administrative and technical metadata, authentication and authorization across federation borders, asynchronous data transfer methods, as well as integrating the technical infrastructure with repositories and the publication systems.

## Ein offener Hafen für Forschungsdaten

In der Wissenschaft liefern große Experimente und Simulationen immer mehr Daten. Um diese Datenmengen dauerhaft, sicher und nachnutzbar speichern zu können, übernimmt nun die Helmholtz-Gemeinschaft die Vorreiterrolle. Sie hat dazu die Helmholtz Data Federation (HDF) gegründet, die durch das SCC am KIT koordiniert wird.

*Kosta Schinarakis<sup>1</sup>*

Die HDF ist ein zentraler Baustein des jüngst verabschiedeten Positionspapiers der Helmholtz-Gemeinschaft zum Umgang mit Forschungsdaten mit dem Titel „Die Ressource Information besser nutzbar machen!“. Durch den sicheren Zusammenschluss (Föderation) und Ausbau von multithematischen Datenzentren ermöglicht es die HDF datenintensiven Wissenschafts-Communities, wissenschaftliche Daten sichtbar zu machen, unter Beibehaltung der Datenhoheit zu teilen, über Disziplinengrenzen hinweg zu nutzen und zuverlässig zu archivieren. Die Föderation basiert auf drei Kernelementen: innovative Software für das Forschungsdatenmanagement, exzellente Benutzerunterstützung und topaktuelle Speicher- und Analyse-Hardware. Dazu werden mittelfristig Investitionen bei den Partnern in Speicher im zweistelligen Petabyte-Bereich und in zehntausende Computerkerne für Datenanalyse und -management durchgeführt. Dafür sind aus Mitteln der strategischen Ausbauinvestitionen in der Helmholtz-Gemeinschaft bis 2021 derzeit 49,5 Millionen Euro eingeplant.

### An Open Haven for research data

Experiments and simulations produce more and more scientific data. The path from the data and information to knowledge needs storage and analysis capabilities of a new quality. The Helmholtz Association takes a pioneering role in securely storing and providing practical access to large amounts of scientific data permanently. For this the Helmholtz Data Federation (HDF) was initiated and is coordinated by the SCC.

„In der Helmholtz Data Federation schließen sich Deutschlands führende Datenzentren zusammen, um die Ströme an Forschungsdaten aus den verschiedenen Wissenschaftsdisziplinen geordnet zu speichern, miteinander zu verknüpfen und einer gemeinsamen Nutzung zuzuführen“, freut sich Professor Achim Streit, Koordinator der HDF und am KIT Direktor des SCC. „Die HDF könnte so zur Blaupause für die datenintensive Forschung in Deutschland und Europa werden, ein offener Hafen für den Zugang zu und den Umschlag von Forschungsdaten.“

Partner der HDF sind in der ersten Phase sechs Zentren aus fünf Forschungsbereichen der Helmholtz-Gemeinschaft: Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (Erde und Umwelt), Deutsches Elektronen-Synchrotron DESY und GSI Helmholtzzentrum für Schwerionenforschung (beide Materie), Deutsches Krebsforschungszentrum (Gesundheit), Forschungszentrum Jülich und Karlsruher Institut für Technologie (beide Energie, Schlüsseltechnologien, Materie, Erde und Umwelt). Die HDF stellt eine Keimzelle für eine wissenschaftsorganisationsübergreifende nationale Forschungsdateninfrastruktur (NFDI) dar, die offen für Nutzer im gesamten deutschen Wissenschaftssystem ist und durch internationale Kooperationen kompatibel mit der zukünftigen European Open Science Cloud (EOSC) sein wird.

### Abgeschlossene Promotionsarbeit

#### Coordinated Caching for High Performance Calibration using Z → μμ Events of the CMS Experiment

*Dr. Max Fischer, KIT-Fakultät für Physik, 22.7.2016*

Forschung im Rahmen des LHC Experiments wird zu einem großen Teil durch Analysen einzelner Wissenschaftler und Gruppen getragen. Klassisch werden hierfür Nutzeranalysen massiv parallel in lokalen Batch Systemen durchgeführt. Dabei werden Datensätze in der Größenordnung von jeweils mehreren Terabyte über das Netzwerk gelesen. Auf einer modernen Multi-Core CPU reichen jedoch schon wenige Analyseprozesse aus, um die verfügbare Netzwerkbandbreite auszureizen.

Im Rahmen dieser Doktorarbeit wurde deshalb in Kooperation mit IEKP und SCC ein neuer Ansatz zur Datenverarbeitung in Batchsystemen entwickelt. Hierbei wird ein Teil der Analysedaten im Batchsystem auf SSDs lokal bereitgehalten, um das Netzwerk zu entlasten. Im Vergleich zu normalen Caches besteht die Herausforderung darin, Daten und Prozesse kohärent zu verteilen. Da die Prozesse von Nutzern beliebig definiert werden, wurde ein System entwickelt, welches die Datenverteilung dem Bedarf anpasst, und gleichzeitig Analyseprozesse optimal platziert.

Seit Sommer 2015 nutzt das IEKP einen Prototyp für Analysen der CMS Kollaboration. Durch das Aufrüsten bestehender Rechenknoten mit SSDs konnte der Durchsatz von Analysen bis auf das 4-fache gesteigert werden.

<sup>1</sup> Presse, Kommunikation und Marketing, KIT

## Laborbuch 2.0 – digital und dezentral verfügbar

Durch die Entwicklung eines dezentralen elektronischen Laborbuchs nutzen Forscher und Wissenschaftler am KIT zukünftig auch in der Labordokumentation viele Vorteile der Digitalisierung. *Jan Potthoff, Nicole Jung<sup>1</sup>*

In den meisten Laboren werden Informationen wie Messergebnisse, Experimentkonfigurationen, Besonderheiten und Bemerkungen zu wissenschaftlichen Untersuchungen häufig handschriftlich aufgenommen - mit allen damit verbundenen Nachteilen: Schlechte Durchsuchbarkeit, hybride Datenhaltung und dazu führender Datenverlust.

Ziel des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekts Entwicklung eines dezentralen elektronischen Laborbuchs ist es, Forschern ein optimal auf deren Bedürfnisse zugeschnittenes Open Source Laborjournal zur Verfügung zu stellen. Dieses Electronic Laboratory Notebook (ELN) wird zunächst auf die Anforderungen von Chemikerinnen und Chemikern ausgerichtet, in weiteren Schritten werden ebenfalls Module zur Nutzung im Bereich Biologie eingebunden. Das ELN wird durch ein Repository ergänzt, welches das Speichern und Veröffentlichen von chemischen Informationen ermöglicht.

Das neuartige Laborbuch soll bisherige Mängel in der Datenerhebung, sowie in der Sicherung analytischer und synthetischer Daten in akademischen Laboren beheben. Darüber hinaus werden neben der traditionellen Publikationsform neuartige Methoden für die Bereitstellung von Informationen etabliert. Im Vordergrund stehen dabei die Verfügbarkeit der Daten für alle Forscher, aber auch die Funktionalität einer geeigneten Langzeitspeicherung. Das Projekt wird gemeinsam vom Institut für organische Chemie (IOC), der KIT-Bibliothek und dem Steinbuch Centre for Computing (SCC) bearbeitet.

Die aktuelle Version des ELN befindet sich in einer Testphase und wird innerhalb der Forschergruppe von Professor Stefan Bräse (KIT/IOC) verwendet. Für interessierte Forscher des Fachbereichs Chemie und angrenzender Disziplinen besteht die Möglichkeit, das ELN auf Nachfrage zu nutzen und an dessen weiterer Gestaltung mitzuwirken.

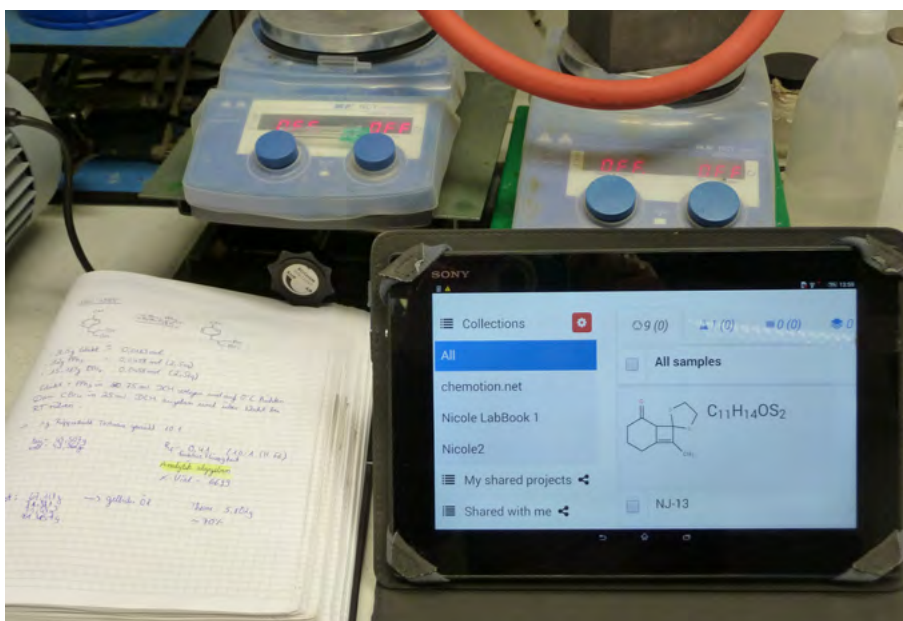
### Laboratory Notebook 2.0 – a digital and decentralized solution

Recording of measurement data, experiment configurations, notes etc. during scientific experiments is still commonly done manually. This practice has several drawbacks e.g. limited searchability, hybrid storage and the subsequent loss of valuable information.

The project 'development of a distributed electronic laboratory journal' develops an open source electronic laboratory journal that aims to fulfill the requirements of scientists working in the laboratory.

The novel electronic journal will fix existing deficiencies with data recording as well as with the protection of analytic and synthetic data produced in scientific laboratories.

Additionally, the project will deliver novel methods for publication and sharing of data directly from the journal.



Chemotion: Das neue elektronische Laborbuch löst die herkömmliche Labordokumentation auf Papier ab.

<sup>1</sup> Institut für Organische Chemie am KIT

## Sternenstaub, Wasserflöhe und Hochleistungsrechner: 3. bwHPC-Symposium am Universitätsrechenzentrum Heidelberg

Was haben die Entstehung von Sternen, die Produktion von OLED-Displays und die Populationsbildung von Wasserflöhen gemeinsam? All diese Prozesse werden mit Hilfe von Hochleistungsrechnern der Rechenzentren in Baden-Württemberg erforscht. Am 12. Oktober 2016 war das Universitätsrechenzentrum (URZ) Heidelberg Gastgeber des inzwischen dritten bwHPC-Symposiums, das über vielfältige Vorträge von Forscherinnen und Forschern einen Eindruck davon vermittelte, für welche unterschiedlichen Zwecke sie das „high performance computing“-Angebot im Land einsetzen.

*Andree Müller<sup>1</sup>, Dr. Robert Barthel*

„High performance computing“ oder kurz HPC wird vor allem dort gebraucht, wo mathematische Modelle numerisch gelöst und große Datenmengen analysiert werden müssen. Das sind Aufgaben, von denen ein normaler Arbeitsplatzrechner im Regelfall überfordert wäre. Mit genügend Rechenleistung lässt sich jedoch die Interaktion physikalischer Kräfte bei der Bildung von Sternen ebenso präzise berechnen, wie die Anordnung einzelner Moleküle bei High-Tech-Produktionsprozessen. Die Vorträge der HPC-Nutzerinnen und Nutzer; im Rahmen des diesjährigen Symposiums; machten erneut deutlich, dass Hochleistungsrechnen in den unterschiedlichsten Fachdisziplinen benötigt und eingesetzt wird. Die Fächer Vielfalt der Beiträge reichte von der Physik und Chemie über die Ingenieurs-, Lebens- und Geowissenschaften bis hin zu den Wirtschafts- und Sozialwissenschaften (s. auch Tab. 1).

Dass Wissenschaftlerinnen und Wissenschaftler einen zuverlässigen und bequemen Zugang zu allen Hochleistungsrechnern haben, ermöglicht das landesweite

Umsetzungskonzept bwHPC<sup>2</sup>. Durch bwHPC wird das vorhandene HPC-Angebot im Land Baden-Württemberg neu organisiert, strukturell vereinheitlicht und vernetzt. Das dazugehörige Begleitprojekt bwHPC-C5 bietet allen HPC-Nutzerinnen und -Nutzern in Baden-Württemberg u.a. vielfältige, fachspezifische Hilfestellung und individuelle Beratung. Das vom Projekt veranstaltete bwHPC-Symposium gewährt regelmäßig Einblicke in Forschungsprojekte, die durch diese landesweite Kooperation unterstützt werden. Das Symposium eröffnet den Dienstbetreibern gleichzeitig die Möglichkeit, den weiteren Aus- und Umbau der HPC-Infrastruktur durch direktes Feedback der HPC-Nutzerinnen und -Nutzer zielführend zu planen.

Die Veranstaltung solle, wie der Direktor des URZ Heidelberg Prof. Dr. Vincent Heuveline betonte, ausdrücklich auch als Ort des Austausches zwischen den Rechenzentren und den HPC-Nutzerinnen und -Nutzern dienen: „Wir wollen hier eine Brücke von der Grundinfrastruktur zu den betrieblichen Aufgaben schlagen.“

Er bedankte sich in seiner Begrüßung besonders beim Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg und auch bei der Deutschen Forschungsgemeinschaft, die durch großzügige Finanzierung sowohl den Erwerb der Hochleistungsrechner als auch die Umsetzung des bwHPC-Konzepts ermöglicht haben.

Das Symposium findet jährlich im Oktober an einer der an bwHPC beteiligten Einrichtungen statt und wurde in den Vorjahren an der Universität Ulm (2015<sup>3</sup>) und am Karlsruher Institut für Technologie (2014<sup>4</sup>) abgehalten.

Mehr Informationen zum bwHPC-Konzept, zum Begleitprojekt bwHPC-C5 und zum Symposium: [www.bwhpc.de](http://www.bwhpc.de)



*Teilnehmerinnen und Teilnehmer des 3. bwHPC-Symposiums*

<sup>1</sup> Universitätsrechenzentrum Heidelberg

<sup>2</sup> Hartenstein, H., Walter, T., Castellaz, P., „Aktuelle Umsetzungskonzepte der Universitäten des Landes Baden-Württemberg für Hochleistungsrechnen und datenintensive Dienste“, PIK – Praxis der Informationsverarbeitung und Kommunikation, Band 36, Heft 2, Seiten 99-108

<sup>3</sup> kiz newsletter, Nov 2015, <http://kiz.uni-ulm.de>

<sup>4</sup> König, T., Barthel, R., „1. bwHPC-Symposium: Erfolgreiche Fortsetzung des Dialogs zwischen Benutzer, Betreiber und Unterstützungsteams“, SCC-News, 02/2014, Seiten 31-33.

<b>Allgemeines</b>	High-Performance Computing and Coordinated Compute Cluster Competence Centers in Baden-Württemberg	Robert Barthel KIT
	The value of and need for system monitoring in HPC operations and research	Florina Ciorba Universität Basel
<b>Physik</b>	Dynamics of the Interstellar Medium and Star Formation	Ralf Klessen Universität Heidelberg
	NEMO's part in investigating the Higgs Boson	Ulrike Schnoor, Felix Bühner Universität Freiburg
<b>Ingenieurwissenschaften</b>	Multiscale Modelling with Accelerated Algorithms	Wolfgang Wenzel KIT
<b>Lebenswissenschaften</b>	CFD Study of the Blood Flow in Cerebral Aneurysms Treated with Flow Diverter Stents	Augusto Sanches Universität Heidelberg
	The genetic population structure of multiple species of Daphnia waterfleas	Robert Kraus Uni Konstanz
	Computational investigation of protein dynamics: exploring the nanoworld through high-performance computing	Davide Mercadante Universität Heidelberg
<b>Wirtschafts- und Sozialwissenschaften</b>	Stochastic Actor-Oriented Models for Network Dynamics	Sebastian Pink Universität Mannheim
	Two-stage stochastic programming of decentralized energy systems	Hannes Schwarz KIT
<b>Chemie</b>	From Experimental Chemistry to 'in silico' Chemistry – Selected Topics of Computational NMR Spectroscopy of Carbocations	Hans-Ullrich Siehl Universität Ulm
	Quantum Chemical Simulations for Astrochemistry	Jan Meisner Universität Stuttgart
	High-level ab initio calculations of molecular anions: what do we need and why do we need it?	Shachar Klaiman Universität Heidelberg
<b>Geowissenschaften</b>	Multi-scale WRF simulations for atmospheric process understanding and boundary layer research	Hans-Stefan Bauer Universität Hohenheim

Tab. 1: Vortrags- und Diskussionsprogramm des 3. bwHPC-Symposiums

### 3<sup>rd</sup> bwHPC Symposium at the Computing Centre of Heidelberg University

The third bwHPC Symposium took place at the Computing Centre of Heidelberg University on October 12, 2016. bwHPC is the current state-wide implementation concept to pool the high performance computing resources and services available at the universities in the state of Baden-Wuerttemberg, Germany. Each year, the bwHPC symposium serves as a discussion board between the HPC service providers and its users. Researchers from various disciplines showed how they make use of the HPC resources and engaged in a lively and constructive discussion with representatives of the facilities running the HPC services.

## „Erste Hilfe“ für Erstsemester

Im Rahmen der SCC-Erstsemestereinführung für das Semester 2016/2017 erhalten die neuen Studierenden einen Überblick über die zentralen IT-Dienste am KIT. Die Informationstage sind auch in diesem Jahr sehr gut bei den Studierenden angekommen. An den insgesamt 18 Vorträgen und 27 Führungen durch das Rechenzentrum haben 670 Studis und die dazu gehörenden Tutoren teilgenommen. Die Zahl der Teilnehmer ist im Vergleich zu den Vorjahren deutlich angestiegen.

*Sabine Grindler*

Weil in den ersten Tagen und Wochen an einem Campus so ziemlich alles neu ist, müssen sich Studienanfänger erst einmal auf dem Campus und in den vielen Gebäuden orientieren und müssen ebenso die administrativen Abläufe für das Studium kennenlernen.

In halbstündigen Vorträgen versorgt das SCC die jungen Studierenden, „Erstis“ genannt, mit allen Infos rund um die zentralen IT-Dienste. Damit die Studis diese Dienste nutzen können, müssen sie zuerst ihr Benutzer-Konto aktivieren. Wie das funktioniert erklären Mitarbeiter des SCC in Kurzvorträgen und geben Antworten auf die Fragen, „Wie kann ich mein E-Mail-Postfach konfigurieren?“,

„Was kostet die Nutzung der zentralen Drucker?“ und „Wie erhalte ich Zugang zu den Campussystemen, wie E-Learning und Prüfungsverwaltung?“ Des Weiteren werden die Poolräume vorgestellt, in denen 400 PCs mit Standard- und Spezialsoftware für die Studis zur Verfügung stehen. In einer anschließenden Führung durch das Rechenzentrum erhielten die Studentinnen und Studenten einen Einblick in die Technik und die vorhandene IT-Infrastruktur.

Zum Nachlesen sind alle Informationen für die neuen Einsteiger im zusammengestellten Flyer „Erste Schritte am KIT“, der auch in die Erstsemesterrucksäcke eingelegt wird.

Unter [www.scc.kit.edu/dienste/flyer](http://www.scc.kit.edu/dienste/flyer) finden Erstsemester alle Infos, die sie in den ersten Wochen am Campus benötigen, angefangen beim Campusplan über die Vorstellung der Einrichtungen und ihrer Öffnungszeiten bis hin zu einem Überblick zu IT-Systemen und Abläufen zur Prüfungsanmeldung.

<http://www.scc.kit.edu/dienste/flyer.php>



Studienanfänger informieren sich über zentrale IT-Dienste am SCC.



## Internationale GridKa Schule – Data Science on Modern Architectures



Vom 29.08. bis 2.9.2016, dieses Jahr zum 14. Mal, war das Steinbuch Centre for Computing Organisator der internationalen GridKa Schule 2016. 130 Teilnehmer, Tutoren und Sprecher aus 12 Ländern fanden den Weg zum KIT, um sich über neue Technologien auf modernen Rechnerarchitekturen zu informieren und sich weiter zu bilden. Das Programm umfasst Vorträge prominenter Sprecher, sowie Seminare, in denen die Teilnehmer anhand praktischer Übungen die Theorie in die Praxis umsetzen.

*Christoph-Erdmann Pfeiler*

Service for Sciences – Science for Services lautet das Motto des SCC, das die integrative Struktur von IT-Diensten und Forschung & Entwicklung in den angewandten Informationstechnologien unterstreicht. Die GridKa-School ist eine gelungene Umsetzung des Teils Service for Sciences. Hier wird Wissenschaftlern und Studenten, Praktikern und Forschern eine Plattform geboten, um im eigenen Fach, aber auch über Fachgrenzen hinweg Erfahrungen zu sammeln und auszutauschen, neue Techniken kennenzulernen und einzuüben.

Das diesjährige Programmangebot beinhaltete Vorträge und Tutorien in den Themenbereichen Big Data Analysis, Large Scale Data Management, Storage Technologies, Databases (SQL and NoSQL), Cloud Installation and Administration, Automation and Configuration Management, Effective Programming, Multi-Core Computing und GPU Computing. Insgesamt gab es 19 Vorträge im Plenum und 16 Tutorials. Einen Gesamtüberblick über das Programm der 14. GridKa Schule gibt die Webseite [gridka-school.scc.kit.edu/2016/135.php](http://gridka-school.scc.kit.edu/2016/135.php).

Der Abendvortrag mit dem Titel „Automated and Connected Driving from a User Centric Viewpoint“ war gut besucht. Prof. Frank Köster vom Deutschen Zent-

rum für Luft- und Raumfahrt (DLR) stellte den bei der DLR verfolgten Ansatz zum autonomen Steuern eines Kraftfahrzeugs vor. Bei dem dort verfolgten Konzept wird die Informationsgewinnung und Verarbeitung teils von lokal im Fahrzeug aufbereiteten Umgebungsdaten, zum anderen aber auch von stationären Einrichtungen zur Verfügung gestellt.

Ein reger Gedankenaustausch nach den Veranstaltungen wurde an mehreren Abenden durch ein entspanntes, wohl organisiertes Beisammensein gefördert. Wer arbeitet, muss auch essen, daher war die Verpflegung während der Schule (19 kg Kekse, 81 kg Obst, 481 l Kaltgetränke und 120 l Kaffee) ein wesentlicher Faktor, der zum Erfolg beitrug. Der traditionelle Flammkuchenabend mit Live-Musik, die Computing Centre Tour durch den Forschungshochleistungsrechner ForHLR II und die Installation des GridKa Computing Centre sowie das Dinner am Donnerstag waren echte Highlights.

Zahlreiche Partner aus Wissenschaft und Industrie, darunter KSETA, IBM, Amazon Web Services, Cloudera, Hitachi, Hewlett Packard, Physics at the Terascale, trugen maßgeblich zum Erfolg der Schule bei.

Dem Organisationsteam um Dr. Manuel Giffels, das für den reibungslosen Ablauf

der Schule verantwortlich ist, gebührt besonderer Dank.

Die nächste GridKa Schule findet vom 28.8. bis 1.9.2017 statt.

### 14<sup>th</sup> International GridKa School – Data Science on Modern Architectures

Also this year the 14th GridKa School organized by Steinbuch Centre for Computing was held in the last week of August. More than 130 participants, including tutors and speakers from 12 nations found the way to KIT, to facilitate knowledge sharing and information exchange. The program covered by the lectures of experts as well as hands-on sessions provided excellent practical experience on Data Science and Modern Architectures. Service for Sciences - Science for Services this is the slogan of SCC, which points to the integrative structure of IT services and research & development in advanced information technologies. GridKa School implements the first part of this slogan - Service for Science. It provides a basis for the different groups of participants like scientists and students, practitioner and researcher, and helps to gain an experience in own profession and beyond, to learn new techniques and to exchange the knowledge.

## Oktober 2016 – der Cyber Security Month

Der European Cyber Security Month (ECSM) ist ein Aktionsmonat zum Thema Cyber-Sicherheit. Im Oktober 2016 sensibilisierten die EU-Mitgliedstaaten Bürgerinnen und Bürger sowie Unternehmen und Organisationen mit Veranstaltungen und Informationen für die IT-Sicherheit. Das Steinbuch Centre for Computing hat einen, auf Forschungseinrichtungen und Universitäten zugeschnittenen, Beitrag geleistet. *Andreas Lorenz*

Der European Cyber Security Month (ECSM) ist ein Aktionsmonat zum Thema Cyber-Sicherheit, der europaweit im Oktober 2016 stattgefunden hat.

Unter Federführung der europäischen IT-Sicherheitsbehörde ENISA (European Union Agency for Network and Information Security) haben die Mitgliedstaaten der Europäischen Union während des Aktionsmonats verschiedene Veranstaltungen und Informationen angeboten, um Bürgerinnen und Bürger sowie Unternehmen und Organisationen für das Thema Cyber-Sicherheit zu sensibilisieren.

Auch das SCC hat einen Beitrag dazu, speziell für Forschungseinrichtungen und Universitäten, geleistet.

Am 21. Oktober hat das Computernotfallteam des KIT (KIT-CERT) eine Trainer-Schulung „IT-Sicherheit für Administratoren“ durchgeführt. Das KIT-CERT und der IT-Sicherheitsbeauftragte des KIT schulen schon seit 2013 regelmäßig IT-Beauftragte und Administratoren (siehe SCC-News 2/2014, S. 46). Aufgrund der guten Vernetzung der Forschungs- und Lehreinrichtungen wurde bald auch Interesse außerhalb des KIT an der Vermittlung des am KIT umgesetzten Konzepts bekundet.



Activities 2016 Challenge Resources About ECSM Contact

Register your Event

### Training programme – IT security for administrators

Karlsruhe Institute of Technology (KIT)



Training programme „Basics of IT-Security in installation“ for admins and IT managers. The training imparts basic knowledge on central questions of IT security to administrators and IT managers: organisation of IT security in installation, secure handling of passwords and accounts, E-Mail security, safe operation of clients and servers, network security, handling of compromised systems

Activity website: [LIT9](#)

Add Event to Calendar: [vCal](#) | [iCal](#)

#### Activities from Germany

Security awareness and encryption of e-mail communication  
13 Oct 2016 from 09:00 to 09:00

Annual conference "10 years of DsiN"  
27 Oct 2016 from 12:00 to 22:00

Kick-off event of the educational radio broadcast  
24 Nov 2016 from 09:00 to 09:00

Digital regulars' table: "Payment via tablet"  
25 Oct 2016 from 12:00 to 19:00

Cybercrime can affect anyone, at any time! - Tips on prevention regarding PC and Smartphone  
01 Oct 2016 to 31 Oct 2016

Nach dem Modell *train the trainer* hat das SCC Personen aus der Helmholtz-Gemeinschaft, den Universitäten im Land Baden-Württemberg und des Verbands TU9<sup>1</sup> mit dem Arbeitsschwerpunkt IT-Sicherheit eingeladen, die solche Schulungen in ihrer Organisation anbieten oder anbieten wollen.

Neben der Darstellung des Schulungskonzeptes wurden Schulungsunterlagen zur Verfügung gestellt, so dass auch in anderen Einrichtungen ähnliche IT-Sicherheitsschulungen angeboten werden können.

Die 13 Teilnehmer haben sich für die gelungene und gewinnbringende Aktion, *Training programme – IT security for administrators*<sup>2</sup>, explizit bedankt. Das SCC hat mit dieser Aktion die IT-Sicherheit auch über die Grenzen des KIT hinweg gefördert.

„Der Workshop bot genau die erwartete Mischung aus fachlichem Inhalt und konzeptioneller Rahmeninformation. Zusammen mit den zur Verfügung gestellten Materialien wird damit die Umsetzung von IT-Sicherheitsschulungen innerhalb der eigenen Organisation deutlich unterstützt“, sagt Dr. Wolfgang zu Castell – Leiter Abteilung Scientific Computing des Helmholtz-Zentrums München. „Die Veranstaltung hat aus meiner Sicht ihr Ziel nach dem 'Train-the-Trainer'-Prinzip voll erreicht. Dass darüber hinaus auch Wert darauf gelegt wurde, persönliche Kontakte aufzubauen, ist ein zusätzlicher Gewinn, welcher dem Ziel der Erhöhung von IT-Sicherheit in Forschungseinrichtungen sicherlich dienlich ist.“, führt er weiter aus.

Abb. 1: Aktion des Steinbuch Centre for Computing zum diesjährigen European Cyber Security Month

### October 2016 – the Cyber Security Month

In order to raise awareness with citizens, businesses and organizations around cyber security and led by the European IT security agency ENISA (European Union Agency for Network and Information Security), EU member states organized various events about information security during last October's cyber-security-month. Especially for research facilities and universities, SCC contributed with a trainer training course "IT Security for Administrators" by the computer emergency response team of KIT (KIT-CERT).

<sup>1</sup> TU9 – das sind neun führende Technische Universitäten in Deutschland: RWTH Aachen University, TU Berlin, TU Braunschweig, TU Darmstadt, TU Dresden, Leibniz Universität Hannover, Karlsruher Institut für Technologie, TU München, Universität Stuttgart.

<sup>2</sup> <https://cybersecuritymonth.eu/ecsm-countries/germany/training-programme-Ait-security-for-administrators>



## Schülerkolloquium Simulierte Welten am SCC

Im öffentlichen Schülerkolloquium präsentierten die Stipendiatinnen und Stipendiaten die Ergebnisse ihrer Projekte zu Themen aus den Bereichen Strömungsmechanik, Hochenergiephysik und Lebenswissenschaften.

*Dr. Ivan Kondov*



*Schülerinnen und Schüler mit ihren Betreuern aus dem SCC*

Im Wintersemester 2015/2016 hat das Steinbuch Centre for Computing im Rahmen des Projekts „Simulierte Welten“ sechs Förderstipendien an Schülerinnen und Schüler der Oberstufe von vier Gymnasien – Bismarck-Gymnasium, Gymnasium Neureut, Markgrafen-Gymnasium und St.-Dominikus-Gymnasium – in Karlsruhe vergeben. Von Mitarbeitern des SCC betreut, haben die Schülerinnen und Schüler die Grundlagen der Modellierung von Computersimulationen, des Hochleistungsrechnens und des datenintensiven Rechnens anhand von drei Themen aus der aktuellen computergestützten Forschung kennengelernt und in der Praxis umgesetzt.

Die Themenbereiche in diesem Jahr umfassten Simulationen von Luftströmungen, Elementarteilchen und Proteinfaltung. Im öffentlichen Schülerkolloquium am 27. Juni 2016 präsentierten die Stipendiatinnen und Stipendiaten die Ergebnisse ihrer Projekte.

Zum Abschluss überreichte Herr Professor Achim Streit, Direktor des SCC, die Urkunden und es gab eine Führung durch die Rechnerräume des SCC. Das Projekt „Simulierte Welten“ wird vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert.

**Weitere Informationen** zum Projekt befinden sich unter:  
[www.simulierte-welten.de](http://www.simulierte-welten.de).

## 31. International Supercomputing Conference

Die 31. International Supercomputing Conference (ISC'16) fand dieses Jahr vom 19.-23. Juni in Frankfurt statt. Die ISC ist die größte europäische und weltweit die zweitgrößte Konferenz für High Performance Computing (HPC). Das SCC präsentierte unter anderem das neue Hochleistungsrechner-System ForHLR II. Dr. Ole Kirner



Bild 1: Stand des Steinbuch Centre for Computing auf der ISC'16.

Den über 3000 Besuchern aus über 50 Ländern wurde ein weitreichendes Konferenzprogramm mit über 400 Vorträgen aus dem Bereich der neuartigen Entwicklungen von HPC- und Big Data- Technologien und deren Nutzung im wissenschaftlichen und kommerziellen Umfeld angeboten. Im Rahmen der ISC fand auch in diesem Jahr wieder eine HPC-Messeausstellung statt. Neben den großen Hardware-Anbietern waren vor allem die großen europäischen Rechenzentren und verschiedenste Softwarehersteller aus dem öffentlichen und industriellen Bereich vertreten.

Das SCC nahm an der HPC-Ausstellung mit einem eigenen Messestand teil (Bild 1) und war darüber hinaus auch an dem Messestand bwHPC beteiligt (Bild 2). Letzterer Stand präsentierte hierbei die Erfolge des baden-württembergischen Umsetzungskonzepts für das Hochleistungsrechnen an den Hochschulen, genannt bwHPC. Ziele von bwHPC sind u.a. die Bereitstellung föderierter HPC-Ressourcen mit dezidierten Fachbereichsschwerpunkten und die hochschulübergreifende Unterstützung der Nutzer auf diesen Ressourcen.

Durch die breite Präsenz des SCC auf der ISC konnten in zahlreichen persönlichen Gesprächen die Inhalte und strategischen Konzepte des SCC an mehrere hundert Besucher weitergeben werden. Inhaltliche Schwerpunkte waren dabei die Infrastruktur und die wissenschaftliche Forschung aus den Bereichen Supercomputing und Big Data.

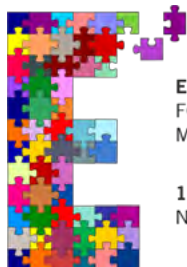
Ein besonderer Schwerpunkt des SCC-Stands war das neue Tier 2 Hochleistungsrechner-System ForHLR II, das im März 2016 am KIT in Betrieb genommen wurde (siehe SCC-News 1/2016, S. 8). Der ForHLR II ist ein Petaflop-System, welches 1170 Knoten, 24.000 Cores, 74 Terabyte Gesamtspeicher und 4.8 Petabyte Plattenplatz umfasst. Der ForHLR kann rechenintensive Simulationen aus verschiedenen wissenschaftlichen Bereichen wie Nanotechnologie, Energie oder Klima und Umwelt durchführen. Durch diese Simulationen werden tiefgreifende Einblicke in aktuelle Forschungsgebiete ermöglicht.

Die Resonanz der Besucher und beteiligten Personen auf unseren Messeständen war auch dieses Jahr wieder sehr positiv. Es ist vorgesehen, dass sich das SCC auch im nächsten Jahr wieder an der ISC beteiligt. Die ISC'17 findet vom 18. bis 22. Juni in Frankfurt statt.



Bild 2: Messestand des Landesprojektes bwHPC.

## Forschungsdaten managen – E-Science-Tage 2017



**E-SCIENCE-TAGE 2017:  
FORSCHUNGSDATEN  
MANAGEN**

**16. - 17. März 2017**  
Neue Universität, Heidelberg

Die Möglichkeiten des digitalen Forschens beflügeln gleichsam Methodik, Vernetzung und Erkenntnisgewinn in der Wissenschaft. Moderne Informationssysteme fördern die Entwicklung innovativer Forschungsmethoden, unterstützen virtuelle Kollaborationen und erhöhen den Wert von Daten für den Forschungsprozess.

Das vom SCC geführte Projekt bwFDM-Info veranstaltet am 16. und 17. März die E-Science-Tage 2017 in Heidelberg. Mit dem Themenschwerpunkt „Forschungsdaten managen“ informiert die Veranstaltung mit einer Fülle von Angeboten über Fortschritte, Herausforderungen und Chancen der digitalen Arbeit mit Forschungsdaten. Die E-Science-Tage bieten sowohl Forschenden als auch Infrastruktureinrichtungen so ein gemeinsames Forum für die Initiierung und Etablierung nachhaltiger Lösungen an.

**Kontakt:** Frank Tristram, frank.tristram@kit.edu

**Weitere Informationen:** [www.e-science-tage.de](http://www.e-science-tage.de)

## Inside SCC – hohes Medien- und Besucherinteresse in 2016

Auch in diesem Jahr hat das SCC seine Türen für zahlreiche Besuchergruppen und Medienvertreter geöffnet. An insgesamt 70 Terminen (37 am Campus Nord, 33 am Campus Süd) wurden über 1300 Besucher durch die Rechner- und Infrastrukturinstallationen geführt.

Im Rahmen der Orientierungsphase für Studienanfänger bekamen 680 junge Menschen, aufgeteilt in 27 Gruppen, einen Einblick in die Rechnerräume am Campus Süd. Dies waren wieder deutlich mehr Studierende als in den Jahren zuvor. Weitere Führungen fanden im Rahmen von Vorlesungen, Schulprojekten, Berufsfindungs- und Studienorientierungstagen, Konferenzen und Festveranstaltungen sowie Kooperationen mit Industriepartnern, Instituten und Arbeitsgruppen am KIT statt.

Neben den unterschiedlichen Besuchergruppen haben viele Medienvertreter den Weg ans SCC gefunden und Foto- und Videoaufnahmen produziert. Einer der Höhepunkte war die festliche Inbetriebnahme des Forschungshochleistungsrechners ForHLR II samt neuem Gebäude und Visualisierungssystem im April 2016. Der Supercomputer hat es unter anderem in den Imagefilm der Stadt Karlsruhe, den Imagefilm der IHK Karlsruhe und sogar in die Tagesschau geschafft. Gemeinsam mit der Fakultät für Informatik wurde speziell für den Studieninformationstag „Uni für Einsteiger“ ein 360°-Video vom Maschinenraum produziert.

Last but not least, konnten die Besucher nicht nur modernste Technik, sondern auch das eigens für den ForHLR-Neubau errichtete Kunstwerk Wachsender Stein von Timm Ulrichs bestaunen und anfassen.

Das SCC bedankt sich bei allen Besuchern und Medienvertretern für das große Interesse und sieht einem spannenden Jahr 2017 entgegen. (ag)

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

#### Redaktion

Achim Grindler (ag, verantwortlich),  
Doris Lang  
E-Mail: [redaktion@scc.kit.edu](mailto:redaktion@scc.kit.edu)

#### Gestaltung, Satz und Layout

Heike Gerstner, Nicole Gross  
PKM – Crossmedia und Marketing  
Grafik  
Karlsruher Institut für Technologie (KIT)  
Hermann-von-Helmholtz-Platz 1  
76344 Eggenstein-Leopoldshafen

#### Titelfotos

Markus Breig, Marcus Hardt,  
Damian Platten, Ellen Rémy

#### Fotos

Robert Barthel, Achim Grindler,  
Marcus Hardt, Andree Müller,  
Olaf Schneider, Achim Streit

#### Druck

Systemedia GmbH, 75449 Wurmberg

**Erscheinungstermin dieser Ausgabe**  
Dezember 2016

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