

## Dirk von Suchodoletz, Janne Chr. Schulz und Jan Leendertse

# Abstraktion erlaubt neue Aufgabenverteilung

Virtualisierung, Clouds und die zukünftige Rolle wissenschaftlicher Rechenzentren



**Technisch möglich, doch noch wenig in Verwendung sind „Virtualisierte Forschungsumgebungen“. Vor dem Hintergrund politischer Vorgaben gibt es erste Blaupausen zu deren Umsetzung. Forschende gewinnen mit Cloud-Technologien räumliche und zeitliche Flexibilität, mit der sowohl Kooperation als auch Eigenständigkeit beim Einsatz von Informationstechnologien neu bestimmbar sind.**

Der große Erfolg der Amazon Webservices (AWS) erklärt sich in der Fähigkeit, Kunden anstehende Investitionen in Hardware und Infrastruktur abzunehmen und durch online buchbare, portionierbare Services zu ersetzen, die ohne großen Vorlauf nutzbar sind (Burrington 2016). Kommerzielle Rechenzentren setzen bei Cloud-Diensten auf Skaleneffekte (Economies of Scale). Nach dem Schritt von der Mainframe auf den allgegenwärtigen PC stellt die Cloud die nächste Iteration der wesentlichen IT-Innovationen dar. Weitere Umwälzungen sind zu erwarten, die heute noch nicht absehbar sind. Obwohl sich im technischen Betrieb Parallelen zwischen AWS und den Prozessen innerhalb der Rechenzentren (RZ) von Hochschulen finden lassen, gibt es betriebliche Anforderungen, mit denen sich Einrichtungen insbesondere in Hochschulen mehr als schwer tun. Planung und Umsetzung von Infrastrukturmaßnahmen im tertiären Bildungsbereich messen sich in Zeiträumen, die eher Jahrzehnte als Jahre umfassen. Das gilt verstärkt für Projekte mit vielen Beteiligten und komplexen Entscheidungsfindungen. Wären Rechenzentren mit der Flexibilität von Start-up-Unternehmen ausgestattet, würden sie ausrufen: „Was für Chancen!“ Sie erfüllen fast alle notwendigen Bedingungen: Sie haben unmittelbaren Zugang zu den Dienstnehmern, sie haben eine berechenbare Einnahmegrundlage, sie haben schon viele Fehler gemacht und damit einen soliden Erfahrungsschatz. Und sie verfügen über passende Services für ihre Mitglieder, die von Studierenden, über Forschende bis in die Administration der Fakultäten reichen. Hier einige Beispiele aus Baden-Württemberg:

- ◆ High-Performance-Computing mit dem Hybrid-Cluster NEMO im landesweiten bwHPC-Verband,
- ◆ Cloud-Lösungen für Angehörige der Universitäten mit bwCloud,

*Die Nutzung von Cloud Technologien im Forschungsalltag ermöglicht eine Alternativlösung zu den bisher genutzten Rechenzentren, die nicht mehr die erforderlichen Leistungen erbringen können.*

*Foto: Pixabay/Geralt*

” **Planung und Umsetzung von Infrastrukturmaßnahmen im tertiären Bildungsbereich messen sich in Zeiträumen, die eher Jahrzehnte als Jahre umfassen.**



Dirk von Suchodoletz leitet seit Juni 2014 die Abteilung eScience am Rechenzentrum der Universität Freiburg. Als Verantwortlicher für HPC und übergreifende Kooperationen von Rechenzentren kennt er Chancen und Risiken aus der Praxis.

Foto: DHBW-Lörrach



Janne Chr. Schulz ist seit Januar 2015 wissenschaftlicher Angestellter am Rechenzentrum der Universität Mannheim. Zurzeit leitet er das Referenzprojekt bw-Cloud mit dem Ziel eines neuen Landesdienstes.

Foto: Timo Hecht



Jan Leendertse ist seit 2015 verantwortlich für die IT-Projektkoordination am Rechenzentrum der Universität Freiburg und leitete dort ein Projekt zur IT-Sicherheit. Zuvor arbeitete er in der freien Wirtschaft im technischen Vertrieb, im Produktmanagement und der Produktentwicklung.

Foto: Helmut Menke

◆ virtuelle Forschungsumgebungen mit digitalen Workflows der Elementarteilchenphysik und Bioinformatik (Meier et al. 2017) und

◆ virtuelle Lehr- und Lernumgebungen mit bwLehrpool (Trahasch et al. 2015).

Der weitere Blick sollte darauf gelenkt werden, was diesen Projekten – damit ist ein Problemfeld schon angerissen – gemeinsam ist:

◆ Sie sind nicht auf eine Hochschule begrenzt,

◆ sie wurden mit Drittmitteln gefördert,

◆ ihre Verstetigung verläuft oft holprig oder im Sande und

◆ sie werden in den beteiligten Hochschulen nicht wirklich als Service wahrgenommen.

Die vorgestellten Dienste befinden sich bislang im Zustand eines Projekts, um trotz der oben genannten Hürden wie der Scheu vor schlecht abschätzbaren Investitionen und Personalrekrutierungen Dinge auszuprobieren und Antworten auf Fragen zu finden, die durch den technologischen Fortschritt an Forschung und Lehre gestellt werden. Der Projektcharakter ist notwendig, weil es keine Erfolgsgarantie gibt und befristete Projekte die notwendige Flexibilität einräumen, Irrwege auszutesten und Erfolgsspuren zu finden, bevor langfristige Stellenhüllen Strukturen fixieren. Virtualisierung und digitale Forschungsprozesse auf Basis von Cloud-Technologien schaffen virtualisierte Forschungs- sowie Lehr- und Lernumgebungen, neue Möglichkeiten des Austauschs, eine Kooperation über institutionelle Grenzen hinweg und die Reproduzierbarkeit von wissenschaftlichen Workflows.

### Die Virtuelle Forschungsumgebung – eine Definition

Eine virtuelle oder virtualisierte Forschungsumgebung (VFU) ist eine bestimmte Form einer lauffähigen Zusammenstellung von Software, in dem je nach Ausprägung das notwendige Betriebssystem und die Applikationen als eigenständig lauffähige Einheit zusammengefasst sind, um wissenschaftliche Aufgabenstellungen zu bearbeiten. Hierzu werden häufig spezielle Konfigurationen an der Software vorgenommen und (externe) Datenquellen und -senken eingebunden. Die VFU verhält sich aus Nutzersicht wie ein eigenständiger Computer, der allerdings nicht direkt auf einem physischen Gerät ausgeführt wird, sondern als eingekapselte Instanz eines weiteren Systems. System-, Prozess- oder Einzelbenutzer werden entweder in der VFU oder durch die vom Hostsystem bereitgestellten Mechanismen authentifiziert und autorisiert. Eine VFU löst somit digitale, wissenschaftliche Methoden und Prozesse aus ihrer fixen Hardware-Umgebung und schafft im Ergebnis eine Beweglichkeit der VFU in Ort und Zeit. Der physikalische Ort, an dem eine Software ausgeführt wird, wird durch schnelle Netze und die Virtualisierung von Rechenleistung und Storage zunehmend unwichtiger. Eine VFU unterstützt in verschiedenen Phasen des wissenschaftlichen Prozesses: Sie ist je nach Disziplin bei der Erhebung, Berechnung oder Verarbeitung von Eingangsdaten, der Simulation von Prozessen bis hin zur Generierung von Ergebnissen einsetzbar. Vielfach sind Werkzeuge für Recherche, Bearbeitung von Texten und Referenzen zur Erstellung der finalen Publikation enthalten. Sie vereinigen das disziplinspezifische Know-how in einem Paket, welches aus Sicht eines Rechenzentrums als Blackbox auf geeigneten Forschungsinfrastrukturen ausgeführt werden kann (Abbildung 1). Durch die Abstraktion der Ebene aus Hardware und Betriebssystem von der Ebene der wissenschaftlichen Werkzeuge können sehr verschiedene Fachdisziplinen ihre Vorhaben und Lehrveranstaltungen durchführen, ohne hierfür weiterhin aufwendig eigene Infrastrukturen zu planen, auszuschreiben, aufzubauen und zu administrieren.

## Stichwörter

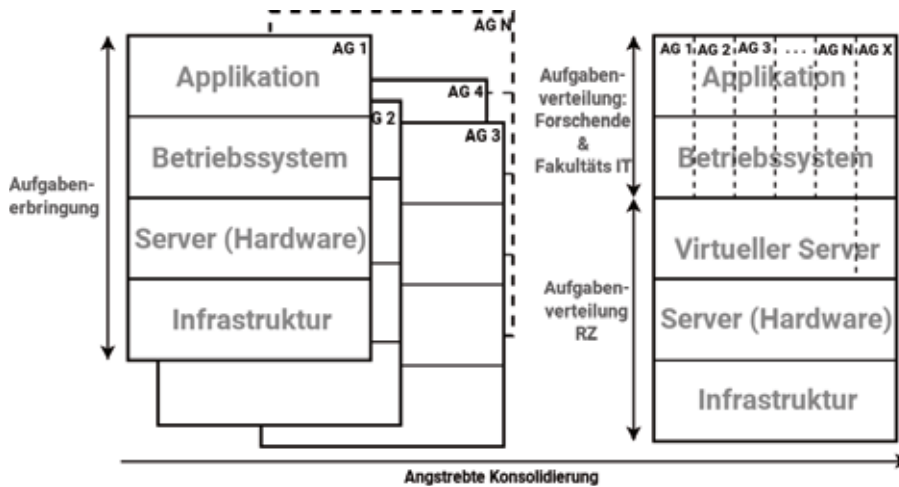
IT-Hochschulstrategie

Serviceabstraktion

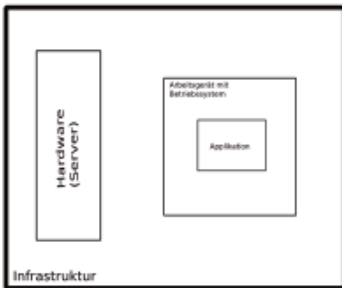
Science-Cloud

Virtualisierte Forschungsumgebungen

Forschungsinfrastrukturen



### Gegenwärtiges Aufgabenspektrum



### Aufgaben in VFU

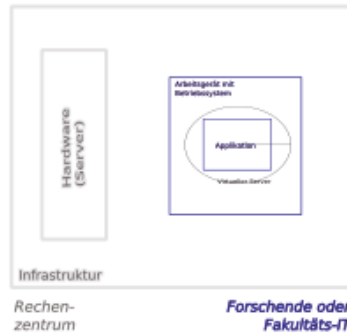


Abb. 1: Verschlanktes Aufgabenspektrum durch Virtualisierte Forschungsumgebungen

Der industrialisierbare Charakter dieser Prozesse vereinfacht die Dokumentation der Arbeit der Forschenden. Zwar lassen sich unbemerkte Veränderungen durch Systemupgrades nicht komplett verhindern, sie sind jedoch zumindest nachvollziehbar dokumentiert und bei Bedarf rückholbar. Läuft eine wissenschaftliche Forschungsumgebung von Beginn an als virtualisierte Instanz, wird die lokale Hardwareabhängigkeit gelockert und späteres Verschieben der Umgebung auf eine neue Virtualisierungsplattform deutlich erleichtert. So könnte ein Wissenschaftler eine Simulationssoftware auf dem lokalen Desktop bereits in einem Container oder einer virtuellen Maschine (VM) entwickeln und in diesem Schritt das Debugging erleichtern, da Algorithmen und Prozesse interaktiv mit direktem Feedback getestet werden können. Die erfolgreich getestete Simulation sollte auf einer Cloud oder in einem Cluster berechnet werden, da der Arbeitsrechner üblicherweise mit der Anzahl der Simulationsläufe überfordert ist. Die hierfür benötigten Ressourcen können durch Rechenzentren in Form einer Compute Cloud bereitgestellt werden. Eine Virtualisierung von Beginn an stellt sicher, dass das Image der VM direkt in die Cloud-Umgebung kopiert und eine oder mehrere für die Problemstellung angemessene virtuelle Instanzen gestartet werden können. Gleichzeitig kann das Image Kooperationspartnern zur direkten Verwendung oder Anpassung an die eigene Fragestellung wie auch Dritten zur Verifikation des Workflows zur Verfügung gestellt oder im Rahmen einer Lehrveranstaltung genutzt werden. Das erleichtert

„Läuft eine wissenschaftliche Forschungsumgebung von Beginn an als virtualisierte Instanz, wird die lokale Hardwareabhängigkeit gelockert und späteres Verschieben der Umgebung auf eine neue Virtualisierungsplattform deutlich erleichtert.“

„Die geeignete Skalierung einer Forschungsinfrastruktur ist für eine wissenschaftliche Arbeitsgruppe oft nur schwer zu beantworten, da sich der tatsächliche Bedarf innerhalb eines Forschungsvorhabens selten exakt vorausbestimmen lässt.“

dem wissenschaftlichen Nachwuchs den Einstieg in die produktive Forschung, da sie nicht mehr ihre Zeit auf die Installation von Betriebssystem und Softwarepaketen ohne Garantie auf Erfolg verwenden. Die geeignete Skalierung einer Forschungsinfrastruktur ist für eine wissenschaftliche Arbeitsgruppe oft nur schwer zu beantworten, da sich der tatsächliche Bedarf innerhalb eines Forschungsvorhabens selten exakt vorausbestimmen lässt. Je nach Fragestellung fließen etliche Variablen in die Größenbestimmung eines Systems ein. Ein Forschungsvorhaben, welches Datensätze von mehreren Gigabyte pro Tag generiert und nach drei Jahren Laufzeit mehrere zehn Terabyte belegt, könnte natürlich ein System in Größe des erwarteten Endausbaus beschaffen. Dieses steht oft erst deutlich nach Projektstart zur Verfügung und ist dann auch noch für den größten Teil der Restzeit deutlich unausgelastet, oder die Arbeitsgruppe muss im Laufe des Projektes Fehlplanungen durch teure Nachbeschaffungen ausgleichen.

### Die Virtuelle Forschungsumgebung – ihre Einbindung und Umsetzung

Rechenzentren unter Anpassungsdruck sowie Rechenzentren in ihren verschiedenen organisatorischen Ausprägungen verspüren zunehmend die Entfremdung von den Entwicklungen in der Wissenschaft. Sie können nicht mehr die vollumfängliche Unterstützung aus der Anfangszeit der IT leisten und sind zunehmend weniger Experten der genutzten wissenschaftlichen Werkzeuge. Dienstnehmer auf dem eigenen Campus empfinden die Angebote des Rechenzentrums als nicht wirklich zu ihren Bedürfnissen passend. Sie schauen sich auf dem kommerziellen Sektor um, der (Forschungs-)Daten wie alle andere Daten als „Zahlungsmittel“ akzeptiert, und finden dort Dienste, die möglicherweise ebenso wenig exakt passen, aber für Forschende günstiger bis kostenfrei und sofort verfügbar sind. Um diesen Trend umzukehren, können VFUs ein Mittel und Ansatzpunkt sein. Sie können damit ihre angebotenen Standarddienste wie Storage oder Server-Hosting zielgruppengerecht aufbereiten, um so eine wirksame Entlastung der einzelnen Fachdisziplinen zu bieten.

Die VFUs bieten je nach erwarteter Aufgabe oder anstehendem Workflow der einzelnen Arbeitsgruppen oder Fachdisziplinen eine skalierende Umgebung, die im Hintergrund auf die bestehenden Basisinfrastrukturen des jeweiligen Rechenzentrums oder zuständigen Verbunddienstes zurückgreift. Hierzu müssen VFUs technisch dem Stand der Entwicklung entsprechen, also die gesamte Palette von der Containerisierung bis zur Virtualisierung beherrschen. Speicher sollte in verschiedenen Formen von schnellem Scratch- bis hochverfügbarem, beziehungsweise redundantem Angebot reichen, der direkt aus der VFU oder vermittelt über das Hostsystem eingebunden werden kann. VFUs bedürfen zur Entfaltung ihrer vollen Wirksamkeit jedoch der Akzeptanz in den jeweiligen Fachwissenschaften. Durch die weitgehend freie Gestaltbarkeit der Inhalte einer VFU entfallen aufwendige Abstimmungsprozesse fast vollständig. Hinzu kommen verkürzte Startzeiten von der Projektidee oder Bewilligung eines Forschungsvorhabens bis zu den ersten Rechen- und Verarbeitungsschritten. Die Forschenden können gemeinsam mit ihren lokalen IT-Admins deutlich fokussierter an ihre Fragestellungen herangehen und sich auf inhaltliche Aspekte konzentrieren. Der direkte und kontinuierliche Kontakt der Rechenzentren zu den Forschenden ist notwendige Voraussetzung, um die Basisangebote von RZs besser zu planen und auf die Anforderungen der Forschenden bestmöglich einzugehen. Sicherstellen lassen sich derartige Kommunikationsstrukturen durch projektbegleitende oder themenbezogene Governancestrukturen. Die so erreichte Standardisierung und Konsolidierung von Infrastruktur am Rechenzentrum entlastet Wissenschaftler von IT-administrativen Aufgaben, die beim Betreiben eigener Infrastrukturen zwangsläufig anfallen. Auf Basis der angestrebten RZ-Infrastruktur ist es möglich, VFUs zu betreiben, die speziell auf die unterschiedlichen Anforderungen der verschiedenen Arbeitsgruppen ausgelegt sind. Die Bereitstellung der Ressourcen muss dabei flexibel

„Rechenzentren unter Anpassungsdruck sowie Rechenzentren in ihren verschiedenen organisatorischen Ausprägungen verspüren zunehmend die Entfremdung von den Entwicklungen in der Wissenschaft.“

und zeitnah erfolgen können. Um die notwendigen Abstimmungen und Entwicklungen gemeinsam durch Wissenschaft und dem Rechenzentrum als Infrastrukturanbieter voranzutreiben, sind sowohl gewisse Anpassungszeiten als auch personelle Ressourcen zu planen. Für den verstärkten Einstieg in virtuelle Forschungsumgebungen wurde daher das vom Land Baden-Württemberg im Rahmen der eScience-Initiative geförderte Projekt „ViCE – Virtual Open Collaboration Environment“ ins Leben gerufen.

### Das Projekt ViCE

Es entwickelt nachhaltige Geschäfts- und Steuerungsmodelle für die Kooperation von unterschiedlichsten Fachdisziplinen mit Rechenzentren auf Basis von VFUs, die dem dynamischen Charakter der Wissenschaften und ihren wechselnden Anforderungen angepasst sind. Es schafft hierzu eine RZ-übergreifende Kollaborations- und Austauschplattform für virtualisierte Forschungsumgebungen, die versioniert, annotiert und geteilt werden können. Die Beschreibung der enthaltenen Tools und Workflows erlaubt die einfache Nachnutzung für neue Forschungsfragestellungen, eine schnelle Einbindung des wissenschaftlichen Nachwuchses und den Einsatz in der Lehre. Hierzu wird Know-how aufgebaut, das die Forschenden in ihren Bedürfnissen unterstützt und eine einfache Ausdehnung auf weitere Fachwissenschaften erlaubt. Die Zusammenarbeit in ViCE erfolgt sowohl über die Grenzen einzelner Fachwissenschaften als auch landesweit über RZ-Grenzen hinweg. Einzelne Basisinfrastrukturen, wie beispielsweise bwCloud, werden nicht mehr von einzelnen RZs, sondern gemeinsam erbracht. Ebenso stellt man fest, dass einzelne Disziplinen wie beispielsweise die Bioinformatik in Freiburg einrichtungsübergreifend bereits seit längerem Dienste wie Galaxy anbietet, die sich für eine Umsetzung als VFU eignen. Ebenso haben verschiedene Fachwissenschaften Fragestellungen gemeinsam wie beispielsweise die nach einer übergreifenden Authentifizierung. ViCE hilft hierbei die Entwicklungen sowohl im Sinne der Rechenzentren als auch der Forschenden voranzutreiben und den Wandel zu moderieren.

### Fazit

VFU sind ein zentrales Mittel, allen Wissenschaftsdisziplinen – Geisteswissenschaften als bislang weniger IT-affine Fachrichtungen eingeschlossen – neue Formen wissenschaftlicher Arbeit und Kooperation zu ermöglichen. VFUs sind zur gleichen Zeit ein Beleg für den fortschreitenden Transformationsprozess der Rechenzentren an Hochschulen als zentrale Einrichtung für alle Angehörigen ihrer Einrichtung. Neben Basisdienstleistungen, beispielhaft seien Vernetzung, Kommunikation und Maschinensäle genannt, sowie aufsetzende Dienste für Campusmanagement, Medienversorgung und elektronischer Lehre, benötigt die Forschung eine neue Form der Unterstützung. Projekte wie ViCE weisen einen Weg, wie sich der Einführungsprozess neuartiger Dienste für einzelne Wissenschaftler oder Forschergruppen moderieren ließe. Sie zeigen, wie Forschung bislang durch IT gesetzte Beschränkungen in Zeit und Raum überwinden und neue Formen der Arbeitsteilung und Kooperation finden können. Notwendige Basisinfrastrukturen der Rechenzentren werden so aufbereitet, dass sie abstrakt von verschiedenen Disziplinen einfach und ohne Startverzögerung eingebunden werden können. Eine zukünftige IT-Strategie moderner Hochschulrechenzentren adressiert neben der Grundversorgung ein aufgefächertes Spektrum an Forschungsinfrastrukturen für Fachwissenschaften. Bei stärkerer Nachfrage werden Skaleneffekte (Economies of Scale) wirken. Diese Neuausrichtung ist längst nicht abgeschlossen und wird anhalten, so lange es gesellschaftliche, politische, hochschulspezifische und technische Veränderungsprozesse gibt. Rechenzentren müssen sich diesen Herausforderungen stellen und Weiterentwicklungen meistern: „Ohne ständige Innovationen an einem IZ/RZ wird es obsolet“ (Wesner 2016, 41).

#### Literatur:

Burrington, I. (2016): Why Amazon's Data Centers Are Hidden in Spy Country. The company powers much of the Internet, but its cloud facilities are difficult to find. URL: <https://www.theatlantic.com/technology/archive/2016/01/amazon-web-services-data-center/423147/> (abgerufen am: 20.10.2017).

Meier, K. et al. (2017): Virtualisierte wissenschaftliche Forschungsumgebungen und die zukünftige Rolle der Rechenzentren, 10. DFN-Forum Kommunikationstechnologien, 30.-31. Mai 2017, Gesellschaft für Informatik e.V. (GI), 145 -154.

Suchodoletz, D. et al. (2016): Kooperation von Rechenzentren. Governance und Steuerung – Organisation, Rechtsgrundlagen, Politik. Berlin, Boston.

Trahasch S. et al. (2015): bwLehrpool: Plattform für die effiziente Bereitstellung von Lehr- und Klausurumgebungen. In: Pongratz, H./Keil, R. (Hrsg): Vol. 247. LNI, DeLFI 2015 – Die 13. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI), München, 1.-4. September 2015 (pp. 291–297). GI., 291-297.

Wesner, S. (2016): Keynote auf der 17. ZKI-Herbsttagung. URL: [https://www.uni-ulm.de/fileadmin/website\\_uni\\_ulm/zkiherbst2016/praesentationen/dienstag/wesner\\_gestaltung\\_von\\_innovationsprozessen\\_an\\_rechen\\_und\\_informationszentren.pdf](https://www.uni-ulm.de/fileadmin/website_uni_ulm/zkiherbst2016/praesentationen/dienstag/wesner_gestaltung_von_innovationsprozessen_an_rechen_und_informationszentren.pdf) (abgerufen am: 15.11.2017).

#### Kontakt:

Dirk von Suchodoletz  
Rechenzentrum  
Universität Freiburg  
Hermann-Herder-Str. 10  
79104 Freiburg i.Br.  
[dirk.von.suchodoletz@rz.uni-freiburg.de](mailto:dirk.von.suchodoletz@rz.uni-freiburg.de)