

# Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS<sup>2</sup>DM)

Gerhard Schneider  Vincent Heuveline Karl-Wilhelm Horstmann  
Bernhard Neumair Petra Hätscher Josef Kolbitsch  Simone Rehm  
Michael Resch Thomas Walter  Stefan Wesner  Peter Castellaz

Das erneuerte Umsetzungskonzept für High Performance Computing wurde um die Bereiche Data Intensive Computing und Large Scale Scientific Data Management erweitert. Hiermit wird eine Erfolgsgeschichte der föderativen Kollaboration der Universitäten des Landes Baden-Württemberg fortgeschrieben. Das neue Konzept adressiert Herausforderungen die sich aus der Digitalisierung der Wissenschaft ergeben. Besonderer Fokus wird auf das nachhaltige Forschungsdatenmanagement gelegt, um den Umgang mit immer größeren Datenmengen zu unterstützen und ihre Nutzbarkeit dauerhaft sicherzustellen. Erklärtes Ziel ist die Etablierung einer landesweiten Datenföderation, um die anstehenden Herausforderungen gemeinsam und kollaborativ adressieren zu können. Komplementiert werden die geplanten Aktivitäten durch eine Erweiterung der Support-Strukturen zur Unterstützung bei Datenhaltungsfragen und einer angepassten Governance, unter Beteiligung von Betreibern und Nutzern.

# 1 Motivation

Computational Sciences (Wissenschaftsrat, 2014) und damit die HPC-Systeme als ihr technisches Fundament gewinnen unablässig an Bedeutung, wie auch der Wissenschaftsrat in seinen jüngsten Empfehlungen zur »Finanzierung des Nationalen Hoch- und Höchstleistungsrechnens in Deutschland« (Wissenschaftsrat, 2015) betont. Die fortschreitende Digitalisierung der Wissenschaft generiert auf Basis verschiedener Forschungsinfrastrukturen Forschungsdaten und damit Anforderungen, die von der schnellen Speicherung bei der Datenerhebung, über die Verarbeitung in HPC- und Cloudsystemen bis hin zur notwendigen Aufarbeitung der Daten im Sinne »guter wissenschaftlicher Praxis« reichen. Die Analyse dieser großen Datenmengen zur Gewinnung von neuen Erkenntnissen wird Data Intensive Computing (DIC) genannt – sie wird heute neben Theorie, Experiment und Simulation als vierte Säule der Wissenschaft bezeichnet. Hinzu kommen die notwendigen technischen und organisatorischen Maßnahmen für eine nachhaltige Nutzung der Daten, die eine langfristige Speicherung und eine nach Möglichkeit öffentliche Zugänglichkeit garantieren.

Der Erkenntnis folgend, dass diese neuen Anforderungen nicht mehr allein und von unabhängig agierenden Akteuren bedient werden können, koordinieren die wissenschaftlichen Rechenzentren des Landes Baden-Württemberg ihre Aktivitäten diesbezüglich. Gleichzeitig wollen die Landesuniversitäten den Empfehlungen des Rats für Informationsinfrastrukturen (RfII) folgen und ihre Infrastrukturentwicklungen mit dem Aufbau einer Infrastruktur für Forschungsdatenmanagement (FDM) auf Basis ihrer HPC- und DATA-Konzepte verschränken. Kooperative Lösungen helfen, die beschriebenen Herausforderungen zu bewältigen, und versprechen einen institutionen- und disziplinübergreifenden Mehrwert.

Für den Zeitraum von 2018 bis 2024 ist es das Ziel aller beteiligten Akteure, den beschrittenen Weg der Kooperation gemäß der HPC-Landesstrategie (MWK-BW, 2017) weiter zu verfolgen. Damit baut das Land Baden-Württemberg ein wesentliches Alleinstellungsmerkmal bei der Unterstützung der Wissenschaften aus und bekundet ausdrücklich das Interesse und die Bereitschaft, in einer frühen Phase beim Aufbau und der Entwicklung der Nationalen Forschungsdateninfrastruktur (NFDI) mitzuwirken. Im Sinne eines integrierten Ansatzes werden die bestehenden Konzepte für HPC, DIC und LS<sup>2</sup>DM weiterentwickelt und in einer gemeinsamen Strategie

zusammengeführt. Gleichzeitig werden die Grundlagen für eine frühe Beteiligung am Aufbau einer NFDI geschaffen und erforderliche Infrastrukturen bereitgestellt.

## 2 Ziele

Die wesentlichen Ziele der geplanten Weiterentwicklung sind die Festigung der für die Forschung und Wissenschaft in Baden-Württemberg bereitgestellten Dienste und Infrastrukturen in den Bereichen HPC, DIC und LS<sup>2</sup>DM und aktive Unterstützung bei Entwicklung und Standardisierung einer NFDI. Weiterhin soll die Einbettung der Landesstrategie HPC, DIC und LS<sup>2</sup>DM in den bundesweiten und europäischen Kontext erfolgen, sowie die Bündelung und Integration der individuellen Aktivitäten in diesem Themenbereich aus den einzelnen wissenschaftlichen Communitys vorangetrieben werden. Ebenfalls geplant sind eine Anpassung der Kooperations-, Governance- und Support-Strukturen sowie der Ausbau und die Schärfung der fachlichen Ausdifferenzierung für qualitativ hochwertige und fundierte Nutzerunterstützung. Das Konzept konkretisiert gleichzeitig das Rahmenkonzept bwDATA (Schneider, Heuveline, Horstmann, Neumair, Waldvogel u. a., 2018) bei der Umsetzung der darin unter »Große wissenschaftliche Daten und Kopplung bwDATA und bwHPC« beschriebenen Vorhaben mit der Erarbeitung der notwendigen konkreten Schritte zur Verknüpfung von HPC- und Forschungsdatenzentren im Land.

Daraus ergeben sich Maßnahmen, die von der Erweiterung der Aufgabenbereiche der Kompetenzzentren um Fragen der Datenhaltung, der Einrichtung weiterer Kompetenzzentren, den bedarfsgerechter Ausbau und Erneuerung der HPC-Systeme aller drei Ebenen und Datenmanagementsysteme, bis hin zur Entwicklung einer landesweiten Datenföderation reichen. Gleichzeitig erfolgt die Fortführung des Begleitprojektes (Barthel u. a., 2019) mit angepassten Schwerpunkten und Ausstattung im Hinblick auf die neuen Anforderungen einer Integration von HPC, DIC und LS<sup>2</sup>DM.

## 3 Entwicklung der HPC-Systeme der Ebenen 1 und 2

Das HLRS wird die Weiterführung des baden-württembergischen Engagements für Tier-0 und Tier-1 in GCS und PRACE verfolgen und daher mit seiner Fokussierung auf Ingenieurwissenschaften eine im Rahmen der deutschen Profilbildung optimale

Architektur für die Forschungsgebiete Mobilität, Energie, Umwelt und Gesundheit umsetzen, um in diesem Bereich seine europaweit führende Stellung auszubauen. Die fortschreitende Einbindung der Simulation in den wissenschaftlichen Forschungsprozess wird darüber hinaus eine entsprechende Integration des eigentlichen Supercomputers in unterstützende Systeme der Datenhaltung und Visualisierung erfordern.

Das SCC übernimmt die Weiterentwicklung der Tier-2-Versorgung und die Unterstützung von Data Analytics. Die anwendungswissenschaftlichen und methodenwissenschaftlichen Schwerpunkte (Materialwissenschaften und Werkstofftechnik, Erdsystem- und Umweltwissenschaften sowie Energie- und Mobilitätsforschung, datenintensives Rechnen) orientieren sich eng an der Profilierung des SCC innerhalb der Gauß-Allianz.

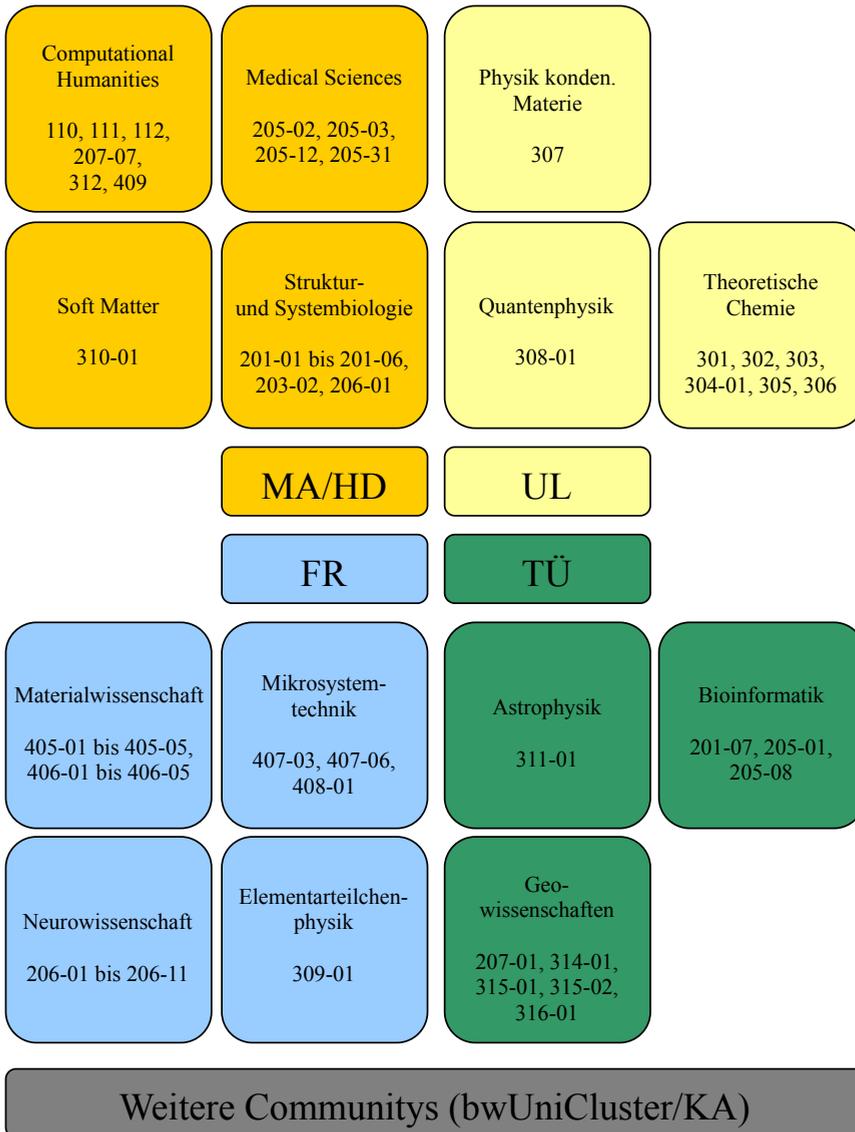
## **4 Entwicklung der HPC-Systeme der Ebene 3**

Die Weiterentwicklung der bwForCluster bedingt bei einer Erneuerung der Systeme die disziplinspezifische Ausrichtung sowohl auf Tier-3 Capacity Computing als auch auf Data Intensive Computing. Damit verbunden ist die Integration der Tier-3-HPC-Systeme mit den angeschlossenen Datenmanagement-Systemen für ihre jeweilige Community und die Anbindung an die HPC-Ressourcen auf Tier-2 und Tier-0/1.

Die bereits praktizierte enge Abstimmung mit den Fachwissenschaften und deren Vertretungen, insbesondere dem Landesnutzerausschuss (LNA-BW) sowie dem Arbeitskreis der Leiter der wissenschaftlichen Rechenzentren in Baden-Württemberg (ALWR-BW) wird dazu fortgeführt.

Die Zuordnung der Fachdisziplinen zu den bwForClustern auf Tier-3 inklusive der zugeordneten speziell ausgeprägten Datenmanagementsysteme soll kontinuierlich weiterentwickelt werden, um weitere Fachdisziplinen und Wissenschaftler und Wissenschaftlerinnen an die Möglichkeiten des Capacity Computing und Data Intensive Computing heranzuführen und bei der Nutzung unterstützen zu können.

Auf Tier-3 soll die fachliche Zuordnung künftig im Detail auf Basis der DFG-Fachsystematik erfolgen (Abbildung 1). Insbesondere bei Erweiterung und Erneuerung der bwForCluster erfolgt eine die bisherige Nutzung der Systeme berücksichtigende Anpassung der fachlichen Widmung. Die Ingenieurwissenschaften nehmen in dieser Zuordnung aufgrund der in diesem Bereich langjährigen bestehenden Er-



**Abbildung 1:** Künftige fachliche Ausdifferenzierung der HPC-Systeme der Ebene 3.

fahrung in der HPC-Nutzung eine Sonderrolle ein. Neben der Grundversorgung auf dem bwUniCluster wird darüber hinausgehender Bedarf direkt auf den Systemen auf Tier-2 und Tier-1 abgedeckt. Das Kompetenzzentrum Ingenieurwissenschaften wird deshalb auch über Ebenen hinweg ausgerichtet und profitiert dabei von den

Erfahrungen in Karlsruhe und Stuttgart. Die erfolgreiche Rolle des bwUniClusters als allgemeines Versorgungssystem für die Universitäten des Landes wird fortgeschrieben.

Die im Rahmen des Landesprojektes bwCloud aufgebauten Infrastrukturen und Dienste ergänzen mit der landesweiten Compute-Cloud-Infrastruktur die Angebote für Nutzer mit geringerem Rechenbedarf, Bedarf an anderen Betriebsmodellen (z. B. primär interaktive Nutzung, permanent laufende Dienste wie Workflow-Engines, Science-Portale) oder speziellen Betriebsumgebungen.

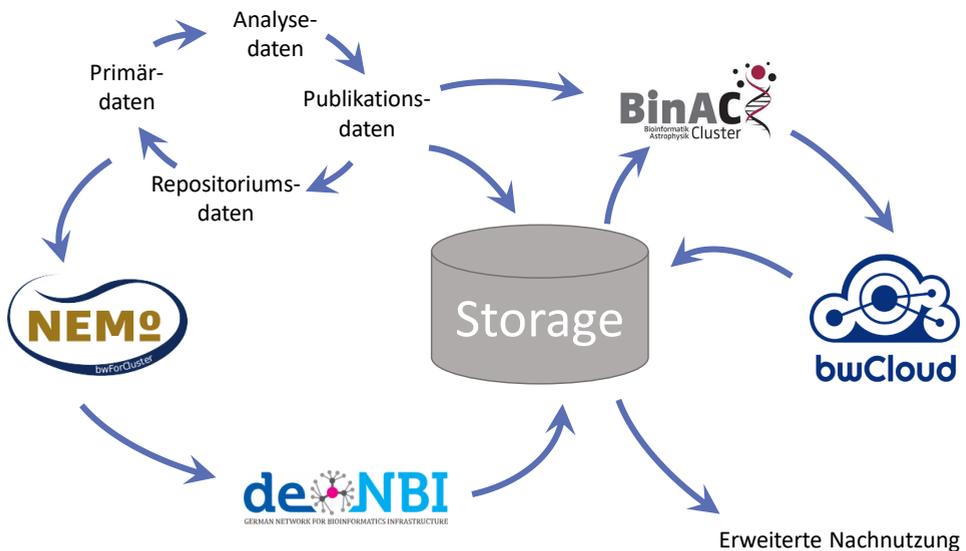
## **5 Data Intensive Computing und Large Scale Scientific Data Management**

Gemäß der Wissenschaftsrats-Empfehlung wird der höchsten Leistungsklasse im HPC (Tier-1) vor allem der Bereich Capability Computing zugeordnet, während die Systeme der nächst niedrigeren Leistungsklasse (Tier-2) sowohl die Anforderungen des Capacity Computing bedienen als auch das Capability Computing adressieren. Die Systeme des Tier-3 dienen im Wesentlichen dem Capacity Computing in den Hochschulen mit Fokus auf individuelle wissenschaftliche Communitys. Baden-Württemberg hat mit seinem bwHPC-Konzept einen leistungsfähigen Rahmen für die Kooperation und Abstimmung zwischen den genannten Ebenen und damit ein sehr gutes Umfeld für die Wissenschaften geschaffen. Diese Grundlagen werden mit dem vorliegenden Konzept weiterentwickelt.

Der zunehmende Einsatz von Methoden zur Datenanalyse, also der Erkenntnisgewinn aus gemessenen oder anderweitig gewonnenen Daten, wurde als aktueller Trend bereits genannt. Häufig sind diese neuen Methoden nicht nur sehr rechenintensiv, sondern sie erfordern zusätzlich auch eine leistungsfähige Verwaltung sehr großer Datenmengen und die Verfügbarkeit großer Datenmengen aus unterschiedlichen Quellen auf HPC-Systemen. Man spricht in diesem Kontext von Data Intensive Computing (DIC).

Ausgehend vom Modell des Data-Life-Cycle lassen sich folgende, typische Workflows und Abläufe identifizieren. Beispielsweise Simulationen auf HPC-Systemen generieren große Datenmengen, die zunächst auf den angeschlossenen, sehr schnellen parallelen Dateisystemen lokal gespeichert werden. Experimente liefern ebenfalls große Datenmengen, die meist nach einer Vorselektion in sogenannte Large Scale

Data Facilities (LSDF) transferiert werden. Die so gewonnenen Daten werden zu einem späteren Zeitpunkt auf speziell dafür ausgelegten Systemen weiter analysiert oder visualisiert. Diese sind häufig aus Performancegründen mit eigenen Speichersystemen ausgestattet. Dies bedingt entweder eine Übertragung großer Datenvolumina von parallelen Dateisystemen oder LSDFs zu diesen Speichersystemen oder aber einen hoch performanten direkten Zugriff auf die jeweilige LSDF. Die erzielten Ergebnisse werden veröffentlicht, und die zugrundeliegenden Daten werden zur Nachnutzung in Repositorien abgelegt. Dies erfordert erneut eine Übertragung und Verarbeitung von Daten auf geeignete Zielsysteme. Am Ende eines Forschungsvorhabens werden die gesammelten Daten für eine spätere erneute Nutzung archiviert und dazu an ein für langfristige Archivierungszwecke ausgelegtes Datenspeichersystem übertragen. Ein Data-Life-Cycle am Beispiel eines Bioinformatik-Workflows ist in Abbildung 2 dargestellt.



**Abbildung 2:** Data-Life-Cycle am Beispiel eines Bioinformatik-Workflows.

Bei all diesen Datenübertragungen müssen Informationen zur Provenienz der Daten, wissenschaftliche Metadaten, Autorisierungs- und ggf. auch Authentifizierungsinformationen mit übermittelt werden. Teilweise sind zusätzlich auch Formatkonversionen oder weitere Verarbeitungsschritte wie Aggregation oder Anonymisierung erforderlich. Die zugrundeliegenden technischen Prozesse können (und sollen) in vielen

Fällen nicht vollständig automatisiert ablaufen, da entsprechende Standards, sofern vorhanden, sehr spezifisch für individuelle wissenschaftliche Communitys sind. Zudem generiert die Interaktion der Wissenschaftler mit den wissenschaftlichen Daten durch Qualitätssicherung und Anreicherung der Metadaten echten Mehrwert.

Ausgehend von den bisher gemachten Erfahrungen beim Betrieb der Dienste im Bereich der Simulation und Daten, soll die bisherige vorwiegend technische Integration auf Basis des föderierten Identitätsmanagement weiter ausgebaut werden. Dabei wird neben einer engeren technischen Verzahnung von Rechen- und Datendiensten insbesondere das Ziel verfolgt, die verschiedenen Dienste und deren Supportstrukturen für die wissenschaftlichen Nutzer in abgestimmten Prozessen und unter einheitlicher Governance anzubieten.

## **6 Forschungsdatenmanagement und Datenföderation**

Das Land Baden-Württemberg hat im Rahmen des gemeinsam entwickelten Fachkonzeptes »E-Science: Wissenschaft unter neuen Rahmenbedingungen – Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg« (MWK-BW, 2014) die Bedeutung des Forschungsdatenmanagements für eine nachhaltige Nutzung von Forschungsdaten betont und zwei Förderprogramme (Forschungsdatenmanagement und Virtualisierte Forschungsumgebungen) eingerichtet. FDM, die nachhaltige und zukunftsorientierte Organisation von Forschungsdaten, ihre Verfügbarmachung und Publikation sind heute ein Fokus moderner Forschungsprozesse. Die Werkzeuge des FDMs sind insbesondere notwendig, wenn die Vorgaben individueller Datenmanagementpläne (DMP) umgesetzt werden müssen und dazu ein Metadatenmanagement erfordern.

Die im Rahmen des vorgelegten Umsetzungskonzeptes zu entwickelnden Maßnahmen werden den Empfehlungen des Rats für Informationsinfrastrukturen (RfII) folgen, indem sie die Weiterentwicklung vorhandener Infrastrukturen mit dem Aufbau einer Umgebung für FDM verschränken. So sollen im Rahmen des Gesamtkonzeptes die beantragten Großgeräte (Erweiterung der LSDF in Heidelberg, bwSFS in Freiburg, Tübingen und Ulm) den Forschern nicht nur große Speichersysteme zur Verfügung stellen, sondern ihnen mittelfristig auch die notwendigen Werkzeuge für das FDM integriert anbieten. Ebenfalls sollen diese Speichersysteme in einer Datenföderation koordiniert werden. Bei der Etablierung der Föderation und zur logischen

Verbindung einzelner Datenmanagementsysteme kann auf verschiedene Quellen zurückgegriffen werden: Existierende Initiativen in der Helmholtz Gemeinschaft (Programm Supercomputing & Big Data), Aktivitäten auf nationaler (BMBF, DFG: LIS-Programme), europäischer (EUDAT, Lecarpentier u. a., 2013; Ardestani u. a., 2015, künftig evtl. EOSC) und internationaler Ebene (Research Data Alliance).

Es hat sich gezeigt, dass für den Zugang zu den Speichersystemen eine Vielfalt an Protokollen angeboten werden muss, um den vielfältigen Anforderungen der Wissenschaftler gerecht zu werden. Dazu gehören neben NFS und CIFS/SMB auch eher aus dem Cloud Umfeld kommende Schnittstellen wie S3. Parallel dazu mussten insbesondere zur Kopplung von Daten und HPC Systemen auch sehr leistungsfähige Anbindungen realisiert werden. Die Umsetzung des Zugangs mit geringen Latenzen und hohen Bandbreiten erfordert entweder kurze Distanzen oder spezielle Dienste wie direkte optische Verbindungen im Weitverkehrsnetz.

In vielen Anwendungsbereichen bestehen für Daten komplexe und restriktive Nutzungsregeln, die besonders bei der zentralen Datenspeicherung berücksichtigt werden müssen. Diese Einschränkungen stehen oft in Zusammenhang mit personenbezogenen (z. B. Medizin, Sozialwissenschaften, Psychologie und Mobilitätsforschung) oder geheimen Daten (z. B. Wirtschaftswissenschaften und Ingenieurwissenschaften). In einer Datenföderation müssen daher auch die Möglichkeiten geschaffen werden, lokale Datenquellen performant anzubinden und beim Datentransfer Mechanismen wie Anonymisierung/Pseudonymisierung oder Aggregation zu unterstützen.

Die Basis für diese Datenföderation stellt vorhandene und zusätzlich aufzubauende Infrastruktur von Datenmanagement- und Datenspeichersystemen dar. Dazu gehören die parallelen Dateisysteme der HPC-Systeme, LSDFs, Datenanalyse-Spezialsysteme, Repositorien und Archivierungssysteme wie bwDataArchiv und bwDataDiss. Ebenso gehört dazu neben lokalen Speichersystemen und Repositorien auch die Anbindung von nationalen und internationalen Systemen, die durch die jeweiligen Fachcommunitys betrieben werden. Diese Systeme sollen und können dabei nicht zu einem oder wenigen zentralen Systemen zusammengeführt werden, sondern grundsätzlich eigenständig Teile einer Föderation bleiben. Dieser Ansatz ermöglicht wie bei bwHPC für viele Wissenschaftsbereiche die bewährte Fokussierung der einzelnen Betreiber auf ihre jeweiligen wissenschaftlichen Communitys und ermöglicht bei Bedarf zusätzlich eine erhöhte Datensicherheit durch Georedundanz. Für Daten, die besonderen datenschutzrechtlichen Anforderungen unterliegen wie z. B. in

der Medizin oder den Sozialwissenschaften, müssen auch dezentrale Ansätze verfolgt und zusätzliche Mechanismen wie Pseudonymisierung, Anonymisierung oder Aggregation der Daten vor einen Transfer etabliert werden. Der Austausch und die Möglichkeit zur Korrelation der Daten müssen unter Beibehaltung der Datenhoheit weitgehend automatisiert und hoch performant ablaufen.

## 7 Weiterentwicklung der Support-Strukturen

Im Rahmen des Begleitprojektes bwHPC-S5 werden die notwendigen Absprachen zwischen allen Betreibern von Teilsystemen der Föderation analysiert, formalisiert und anschließend langfristig etabliert. Das Betriebsprojekt wird auch die angesprochenen Fragen notwendiger Formatkonversionen, Abbildungen von Metadaten und von Autorisierungs- und Authentifizierungsinformation adressieren. Die Benutzerunterstützung stellt neben den technischen Systemen für Datenmanagement und Datenspeicherung sowie der Datenmanagement-Software die dritte wesentliche Säule bei der Etablierung der Datenföderation dar.

Kernelement des neu ausgerichteten, erweiterten Begleitprojektes bwHPC-S5 (Scientific Simulation und Storage Support Service) ist die Weiterentwicklung von föderativen Unterstützungsstrukturen. Bei der Integration der Systeme und des Datenmanagements steht der sichere und performante Zugriff auf entfernte Daten sowie die Integration in den Data-Life-Cycle im Projektfokus, um das datenintensive Rechnen zu ermöglichen. Im Bereich Infrastruktur und föderatives Dienstemanagement wird das Augenmerk auf Monitoring, Reporting und betriebliche Optimierung gelegt. Unverzichtbar ist die Ausweitung des landesweiten HPC-Schulungsprogramms unter Berücksichtigung von DIC und LS<sup>2</sup>DM. Bei allen Projektzielen wird verstärkt auf die Abstimmung mit den Tier-1- und Tier-2-Kompetenzen geachtet.

## 8 Governance

Die wissenschaftsgeleitete Steuerung der Nutzung und Weiterentwicklung dieser Infrastruktur wird gewährleistet durch eine übergreifende und in den letzten Jahren etablierte Governance- Struktur (Wesner, Suchodoletz u. a., 2017; Wesner, Walter u. a., 2016). Sie besteht aus den folgenden Elementen: HLRS-Lenkungsausschuss

(für HPC Tier-1 und Tier-2), Landesnutzerausschuss (für HPC Tier-3), bwHPC-Lenkungskreis (für HPC Tier-3) sowie bwDATA-Steuerungskreis.

Der HLRS-Lenkungsausschuss stellt das seit langem etablierte Aufsichtsgremium für das Tier-1-System in Stuttgart und das Tier-2-System in Karlsruhe dar. Er ist besetzt mit Wissenschaftlerinnen und Wissenschaftlern, die je zur Hälfte auf Vorschlag der DFG und der Landesrektorenkonferenz vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg berufen werden. Er realisiert ein wissenschaftsgeleitetes Zugangsverfahren zu den beiden Systemen und stellt die optimale und offene Nutzung der Ressourcen für die jeweiligen Anwendungen und die Durchlässigkeit zwischen den Systemen sicher.

Für die wissenschaftliche Steuerung der bwHPC-Dienste auf Tier-3 und die nutzerseitige Steuerung im Bereich der wissenschaftlichen Datenhaltung wurde der Landesnutzerausschuss Baden-Württemberg (LNA-BW) eingerichtet. Im LNA-BW sind alle Universitäten des Landes sowie Repräsentanten der Hochschulen für Angewandte Wissenschaften vertreten. Die Prorektoren/Vizepräsidenten für Forschung schlagen jeweils ein Mitglied vor, das dann vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg in den LNA-BW berufen wird. Zudem sind die Hochschulen für Angewandte Wissenschaften mit einem gemeinsamen Vertreter Mitglied im LNA. Die Aufgaben des Landesnutzerausschusses umfassen aktuell die Diskussion der Regularien bezüglich des Ressourcenzugangs, die Beurteilung von Clusterauslastungsdaten sowie die Diskussion der Fachbereichszuordnung der Kompetenzzentren. Darüber hinaus behandelt der Ausschuss die Regulierung von Clustererweiterungen und die Vertretung der Nutzerwünsche bezüglich des operativen Betriebes gegenüber den Betreibern. Dazu gehören die Rückmeldung des nutzerseitigen Bedarfs an neuen Technologien an die Betreiber, die Ermittlung des Bedarfs an Software-Lizenzen, Bedarf an Quotas sowie Anpassungen bei Job-Queues bezüglich Job-Laufzeiten.

Der bwHPC-Lenkungskreis ist besetzt mit Mitgliedern aus dem Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg, dem LNA-BW, dem ALWR und den Hochschulen für angewandte Wissenschaften. Er übernimmt die Rolle des Lenkungsausschusses für die Projekte im Umfeld von bwHPC, insbesondere für das erweiterte Begleitprojekt bwHPC-S5. Der bwHPC-Lenkungskreis begleitet den LNA-BW und den ALWR-BW in allen Fragen betreffend Konzept und Evaluierung von bwHPC, Bedarfserhebungen, Ressourcen, Betrieb und Begleitprojekte.

Der bwDATA-Steuerungskreis überwacht analog zum bwHPC-Lenkungskreis die Umsetzung der im Rahmenkonzept bwDATA festgelegten Projekte und Maßnahmen für große wissenschaftliche Daten. Ihm gehören die Mitglieder des ALWR-BW an sowie mindestens zwei Vertreter der Hochschulen für Angewandte Wissenschaften, ein Vertreter der Dualen Hochschule Baden-Württemberg (DHBW), sowie bei Bedarf je ein Vertreter der Pädagogischen Hochschulen, des Landesarchivs, der Landesbibliotheken, der Kunst- und Musikhochschulen und der vier Universitätskliniken. Zu den zentralen Aufgaben des Steuerkreises gehört auch die Wahrnehmung der Verantwortung für die kontinuierliche konzeptionelle Fortführung des Rahmenkonzeptes bwDATA. An den Sitzungen nimmt ein Vertreter des MWK als ständiger Gast teil.

Im Zuge der Einführung der Datenföderation für Baden-Württemberg soll die Governance ergänzt werden, um die weiteren Aufgaben dieses Konzeptes abzudecken. So ist eine Erweiterung der Aufgaben des Landesnutzerausschusses von den bisher zentral behandelten Fragen des High Performance Computings hin zu datenintensiven Diensten vorgesehen. Darüber hinaus ist eine Zusammenführung von bwHPC-Lenkungskreis und bwDATA-Steuerkreis geplant, um die in diesem Umsetzungskonzept adressierten Herausforderungen effizient zu bewältigen.

**Gekürzte Fassung** Dieser Text ist die gekürzte Fassung des von der DFG begutachteten »Umsetzungskonzeptes der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS<sup>2</sup>DM)« von Schneider, Heuveline, Horstmann, Neumair, Hätscher u. a. (2019).

## Autoren

Gerhard Schneider, Universität Freiburg  
Vincent Heuveline, Universität Heidelberg  
Karl-Wilhelm Horstmann, Universität Hohenheim  
Bernhard Neumair, Karlsruher Institut für Technologie  
Petra Hätscher, Universität Konstanz  
Josef Kolbitsch, Universität Mannheim  
Simone Rehm, Universität Stuttgart  
Michael Resch, Universität Stuttgart  
Thomas Walter, Universität Tübingen

Stefan Wesner, Universität Ulm

Peter Castellaz, Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg

## Korrespondenzautoren

Gerhard Schneider: [gerhard.schneider@rz.uni-freiburg.de](mailto:gerhard.schneider@rz.uni-freiburg.de)

Rechenzentrum, Albert-Ludwigs-Universität, Freiburg

Thomas Walter: [thomas.walter@uni-tuebingen.de](mailto:thomas.walter@uni-tuebingen.de)

ZDV, Eberhard-Karls-Universität, Tübingen

## ORCID

Gerhard Schneider  <https://orcid.org/0000-0002-3214-002X>

Josef Kolbitsch  <https://orcid.org/0000-0002-7601-1553>

Thomas Walter  <https://orcid.org/0000-0002-8656-2340>

Stefan Wesner  <https://orcid.org/0000-0002-7270-7959>

Lizenz  4.0 <https://creativecommons.org/licenses/by-sa/4.0>

## Literatur

Ardestani, S. B. u. a. (2015). »B2share: An open escience data sharing platform«. In: *2015 IEEE 11th International Conference on e-Science (e-Science)*. IEEE, S. 448–453.

Barthel, R. und J. Salk (2019). »bwHPC-S5: Scientific Simulation and Storage Support Services. Unterstützung von Wissenschaft und Forschung beim leistungsstarken und datenintensiven Rechnen sowie großskaligem Forschungsdatenmanagement«. In: *Proceedings of the 5th bwHPC Symposium. HPC Activities in Baden-Württemberg*. Freiburg, September 2018. 5th bwHPC Symposium. Hrsg. von M. Janczyk, D. von Suchodoletz und B. Wiebelt. TLP, Tübingen, S. 17–28. DOI: [10.15496/publikation-29039](https://doi.org/10.15496/publikation-29039).

Lecarpentier, D. u. a. (2013). »EUDAT: a new cross-disciplinary data infrastructure for science«. In: *International Journal of Digital Curation* 8.1, S. 279–287. DOI: [10.2218/ijdc.v8i1.260](https://doi.org/10.2218/ijdc.v8i1.260).

MWK-BW (2014). *E-Science: Wissenschaft unter neuen Rahmenbedingungen. Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg*. Fachkonzept. Ministerium für Wissenschaft, Forschung und Kunst. URL: [https://www.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/Anlagen\\_PM/2014/066\\_PM\\_Anlage\\_E-Science\\_Web.pdf](https://www.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/Anlagen_PM/2014/066_PM_Anlage_E-Science_Web.pdf) (besucht am 25.02.2019).

- MWK-BW (2017). *HPC-Landesstrategie: Eine halbe Milliarde Euro für digitale Infrastruktur und Supercomputer*. Pressemitteilung Nr. 99/2017. Ministerium für Wissenschaft, Forschung und Kunst. URL: [https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/Anlagen\\_PM/2017/099\\_PM\\_Super-Computing\\_Eine\\_halbe\\_Milliarde\\_Euro\\_f%C3%BCr\\_digitale\\_Infrastruktur.pdf](https://mwk.baden-wuerttemberg.de/fileadmin/redaktion/m-mwk/intern/dateien/Anlagen_PM/2017/099_PM_Super-Computing_Eine_halbe_Milliarde_Euro_f%C3%BCr_digitale_Infrastruktur.pdf) (besucht am 07.02.2019).
- Schneider, G., V. Heuveline, K.-W. Horstmann, B. Neumair, P. Hätscher u. a. (2019). *Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS<sup>2</sup>DM)*. Konzept. Hochschulen des Landes Baden-Württemberg. URN: urn:nbn:de:bsz:21-dspace-864846.
- Schneider, G., V. Heuveline, K.-W. Horstmann, B. Neumair, M. Waldvogel u. a. (2018). *Rahmenkonzept der Hochschulen des Landes Baden-Württemberg für datenintensive Dienste – bwDATA (2015-2019)*. de. Konzept. Hochschulen des Landes Baden-Württemberg. DOI: 10.15496/publikation-21187.
- Wesner, S., D. von Suchodoletz, B. Wiebelt, G. Schneider und T. Walter (2017). »Overview on governance structures in bwHPC«. In: *Proceedings of the 3rd bwHPC-Symposium: Heidelberg 2016*. (2016). Hrsg. von S. Richling, M. Baumann und V. Heuveline. Heidelberg: heiBOOKS. DOI: 10.11588/heibooks.308.418.
- Wesner, S., T. Walter, B. Wiebelt, D. von Suchodoletz und G. Schneider (2016). »Strukturen und Gremien einer bwHPC-Governance – Momentaufnahmen und Perspektiven«. In: *Kooperation von Rechenzentren Governance und Steuerung – Organisation, Rechtsgrundlagen, Politik*. Hrsg. von D. von Suchodoletz, J. C. Schulz, J. Leendertse, H. Hotzel und M. Wimmer. de Gruyter, S. 315–329. ISBN: 978-3-11-045888-6. DOI: 10.1515/9783110459753-027.
- Wissenschaftsrat (2014). *Bedeutung und Weiterentwicklung von Simulation in der Wissenschaft*. Positionspapier. WR. URL: <https://www.wissenschaftsrat.de/download/archiv/4032-14.pdf> (besucht am 07.02.2019).
- (2015). *Empfehlungen zur Finanzierung des Nationalen Hoch- und Höchstleistungsrechnens in Deutschland*. Empfehlungen. WR. URL: <https://www.wissenschaftsrat.de/download/archiv/4488-15.pdf> (besucht am 07.02.2019).