

Rahmenkonzept der Universitäten des Landes Baden-Württemberg für das High-Performance Computing (HPC) und Data-Intensive Computing (DIC) für den Zeitraum 2025 bis 2032

Vorgelegt von den Leitern der Rechenzentren bzw. Informationszentren der Universitäten des Landes Baden-Württemberg (ALWR-BW)

Dr. Dirk von Suchodoletz, Universität Freiburg
Prof. Dr. Vincent Heuveline, Universität Heidelberg
Stefan Farrenkopf, Universität Hohenheim
Prof. Dr. Bernhard Neumair, Karlsruher Institut für Technologie
Oliver Kohl-Frey, Universität Konstanz
Alexander Pfister, Universität Mannheim
Jörn Beutner, Universität Stuttgart
Prof. Dr.-Ing. Dr. h.c. mult. Michael Resch, Universität Stuttgart
Prof. Dr. Thomas Walter, Universität Tübingen
Thomas Nau, Universität Ulm
Dr. Raphael Dorn, Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg
(ständiger Gast)

unter Beteiligung von

Prof. Dr. Martin Frank, Karlsruher Institut für Technologie
Prof. Dr. Caroline Ruiner, Universität Hohenheim
Prof. Dr. Gerhard Schneider, Universität Freiburg
Prof. Dr.-Ing. Dr. h.c. Stefan Wesner, Universität Ulm

in Abstimmung mit den Hochschulen für angewandte Wissenschaften des Landes Baden-Württemberg

Revision 1.0, Juni 2023

Inhaltsverzeichnis

INHALTSVERZEICHNIS	2
1 STRATEGISCHE ZIELSETZUNG	3
2 AUSGANGSSITUATION	5
2.1 HPC-LEISTUNGSPYRAMIDE IN BADEN-WÜRTTEMBERG	5
2.1.1 TIER-1.....	5
2.1.2 TIER-2.....	5
2.1.3 TIER-3.....	5
2.2 DATENMANAGEMENT UND SPEICHERSYSTEME	7
2.3 BELWÜ	8
2.4 WISSENSCHAFTSUNTERSTÜTZUNG UND GOVERNANCE	8
2.5 WISSENSTRANSFER / WIRTSCHAFT	9
3 HERAUSFORDERUNGEN UND HANDLUNGSFELDER	11
3.1 WISSENSCHAFTLICHE ANFORDERUNGEN	11
3.2 ENERGIEEFFIZIENZ UND NACHHALTIGKEIT	12
3.3 SYSTEMKOMPONENTEN UND -ARCHITEKTUREN	13
3.4 SOFTWARE	14
3.5 DATENMANAGEMENT UND SPEICHERSYSTEME	16
3.6 UMGANG MIT SENSIBLEN DATEN ETWA AUS DER MEDIZIN	17
3.7 DIGITALE SOUVERÄNITÄT	18
3.8 STRUKTUREN UND ORGANISATION	19
4 QUELLEN UND REFERENZEN	21

1 Strategische Zielsetzung

Digitale Infrastrukturen und darauf aufsetzende Dienste bilden inzwischen in fast allen Fachgebieten das Rückgrat wissenschaftlicher Forschung. Deshalb muss die strategische Zielsetzung eines HPC-Konzepts klare Antworten auf eine Vielzahl von Problemstellungen finden. Dazu zählen die stetige Zunahme digitaler Workflows, bedingt beispielsweise durch die Verbesserung des Instrumentariums in den Naturwissenschaften, neue Forschungsansätze in den Digital Humanities, die Verfeinerung der Auflösung in bildgebenden Verfahren, aber auch der zunehmende Einsatz Künstlicher Intelligenz in immer breiteren Anwendungsfeldern. Öffentlich geförderte und betriebene Forschungsinfrastrukturen sind hier von entscheidender Bedeutung. Föderierte Strukturen erzeugen Synergien und haben klare strategische Vorteile gegenüber vereinzelt Insellösungen, insbesondere im Bereich High-Performance-Computing/Data Intensive Computing (HPC-DIC). Daher müssen sie im Interesse der Nutzenden auf der lange etablierten Basis verlässlich, zukunftssicher und nachhaltig weiterentwickelt werden. Im Sinne des Wissenschafts- und Wirtschaftsstandorts Baden-Württemberg sollen sie eine für Menschen gemachte Digitalisierung voranbringen, moderne technische Entwicklungen mitgestalten, zusätzliche Kreativität freisetzen, und nicht zuletzt die breite Verteilung von Kompetenzen unterstützen. Basierend auf internationalen Standards sind digitale Dienste auf allen Ebenen (lokal, regional, national und international) geeignet zu vernetzen. Um den Standort Baden-Württemberg weiter attraktiv für Forscher, Entwickler und Dienstleister auszubauen, ist dies eine unverzichtbare Voraussetzung. Die Hochschulen im Land müssen sich im Hinblick auf die Anwerbung von IT-Fachkräften zu einem begehrten Arbeitgeber weiterentwickeln [4].

Die übergeordnete strategische Zielsetzung lässt sich zusammenfassen als:

- Gewährleistung einer stets weltweit konkurrenzfähigen Hardware- und Softwareversorgung für die gesamte baden-württembergischen Wissenschaft im Bereich des HPC/DIC
- Integration in die europäische und nationale HPC/DIC-Strategie wie sie sich im Joint Undertaking EuroHPC [5], im Gauss Centre for Supercomputing (GCS) [6], im Nationalen Hochleistungsrechnen (NHR) [7] sowie in der Nationalen Forschungsdateninfrastruktur (NFDI)[5] manifestiert
- Transfer der in den Bereichen HPC/DIC erzielten Erkenntnisse und Fähigkeiten über den wissenschaftlichen Bereich hinaus in Gesellschaft und Wirtschaft

Neben die wissenschaftlichen Herausforderungen des Hoch- und Höchstleistungsrechnens treten zunehmend Herausforderungen im Betrieb der Systeme. Vor dem Hintergrund der Risiken des Klimawandels müssen sich Rechenzentren in allen Belangen noch gezielter an Nachhaltigkeitskriterien ausrichten, zukünftigen gesetzlichen Anforderungen genügen und einen Beitrag zur Reduktion des CO₂-Ausstoßes leisten. Eine zentrale Aufgabenstellung der kommenden Jahre ist daher, eine echte Klimaneutralität von Hoch- und Höchstleistungsrechenzentren zu erreichen. Als besondere Handlungsfelder sind zu nennen:

- Optimierung der Energieeffizienz mit Ziel der Klimaneutralität und Nachweis durch entsprechende Zertifizierung; dazu Untersuchung und Umsetzung neuer Betriebsmodelle und Auslegung der Systeme im Hinblick auf Betrieb mit (volatilen) erneuerbaren Energieträgern; Ertüchtigung von Bestandsgebäuden und weitergehende Standortkonsolidierung mit dem Ziel einer Umsetzung eines klimapositiven Rechenzentrums; Sensibilisierung von Nutzenden
- Verstärkte Hinwendung zur nachhaltigen Entwicklung wissenschaftlicher Software, insbesondere zur Steigerung der Energieeffizienz; verbesserte Qualifizierung des

wissenschaftlichen Nachwuchses, um HPC/DIC-Systeme oder neue Technologien wie KI deutlich effizienter zu nutzen

Über diese allgemeinen Ziele hinaus sieht die Strategie, wie bisher schon, auch für die Jahre 2025 bis 2032 vor, auf aktuelle Herausforderungen flexibel zu reagieren, um wissenschaftliche Neuentwicklungen schnell und zielgerichtet durch die notwendige digitale Infrastruktur unterstützen und fördern zu können. Zu den erwarteten neuen Herausforderungen gehören insbesondere:

- Forschungsdatenmanagement (FDM);
- rapide Zunahme der Forschung im Bereich der Künstlichen Intelligenz und ihre Nutzung in allen Wissenschaftsbereichen;
- Erforschung und Nutzung neuer Hardwarekonzepte im HPC/DIC und hier insbesondere des Quantum Computing (QC).

Um diese gestiegenen Anforderungen zu adressieren, schreiben die Hochschulen Baden-Württembergs mit diesem Papier zum dritten Mal ihr Rahmenkonzept für die Bereiche des Hoch- und Höchstleistungsrechnens (HPC) einerseits und des Data Intensive Computing andererseits fort. Sie tragen damit der weiterhin steigenden Bedeutung von Computational Sciences und HPC-Systemen als deren Fundament Rechnung.

2 Ausgangssituation

Baden-Württemberg hat sich in den letzten 40 Jahren im HPC-DIC Umfeld klar und strategisch positioniert. Schon in den 1980er Jahren wurden an einigen Standorten in Baden-Württemberg HPC-Systeme beschafft und mit dem Wissenschaftsnetz BelWü die notwendige Konnektivität bereitgestellt. Diese Basis wurde seither sukzessive strategisch und technisch weiterentwickelt und mit nationalen und internationalen Entwicklungen verwoben. Dies stärkt den Wissenstransfer innerhalb der Partnernetzwerke aber auch zu Nutzenden und in die Industrie. Externe wissenschaftliche Gutachtende haben im Laufe der letzten Jahre mehrfach herausgestellt, dass sich die in Baden-Württemberg etablierten Lösungen auch als Modell für nationale und internationale Versorgungsstrukturen eignen.

Während der Laufzeit des vorliegenden Rahmenkonzeptes werden sowohl die unter 2.1 beschriebenen HPC-Komponenten als auch die unter 2.2 aufgeführten Speichersysteme kontinuierlich nach den Bedürfnissen erneuert und ausgebaut; beispielsweise wird an der Universität Ulm mit Justus 3 die nächste Stufe dieses bwForClusters aktuell geplant.

2.1 HPC-Leistungspyramide in Baden-Württemberg

2.1.1 Tier-1

Auf europäischer Ebene (Tier-0) übernimmt das Gauss Centre for Supercomputing (GCS) mit den Partnern LRZ (Garching), HLRS (Stuttgart) und JSC (Jülich) die Rolle eines ersten europäischen Exaflop-Providers mit dem Standort Jülich. Die Finanzierung erfolgt im Rahmen des Projekts SiVeGCS/SiVeGCS+ sowie durch die Joint Undertaking EuroHPC. Im Rahmen der GCS-internen Abstimmung übernimmt das HLRS die Aufgabe der Koordinierung großer europäischer Projekte für den Exascale-Bereich von EuroHPC. Aktuell werden die Projekte EuroCC und CASTIEL geleitet.

Auf deutscher Ebene (Tier-1) realisiert das Höchstleistungsrechenzentrum Stuttgart (HLRS) der Universität Stuttgart die Funktion eines nationalen und europäischen Höchstleistungsrechenzentrums. Im Rahmen des vom BMBF und den drei Landesministerien für Wissenschaft geförderten Projekts SiVeGCS wurde am HLRS ein Tier-1 System 2020 in Betrieb genommen. Im Rahmen der 2021 beschlossenen Erweiterung von SiVeGCS+ soll im Jahr 2024 ein Übergangssystem installiert werden und im Jahr 2026/2027 ein Exascale-System am Standort HLRS den Betrieb aufnehmen. Dieses System wird bis 2032 für die nationale Versorgung im Höchstleistungsrechnen zur Verfügung stehen.

2.1.2 Tier-2

Auf Tier-2 übernimmt das Steinbuch Centre for Computing (SCC) am Karlsruher Institut für Technologie (KIT) die Rolle eines Hochleistungsrechenzentrums und Zentrums für datenintensives Rechnen. Im Rahmen des Nationalen Hochleistungsrechnens (NHR) betreibt das SCC das Tier-2-System HoreKa, das insbesondere die Materialwissenschaften, die Erdsystemwissenschaften, die Energie- und Mobilitätsforschung im Ingenieurwesen sowie die Teilchen- und Astroteilchenphysik unterstützen soll. Innerhalb des NHR-Verbunds leistet das SCC im Bereich der Methodenwissenschaften federführend die Unterstützung für das Research Software Engineering und trägt wesentlich zum Bereich numerische Methoden und Bibliotheken bei.

Im betrachteten Zeitraum dieses Papiers sind jeweils zu Beginn und zum Ende hin große Rechnerbeschaffung vorgesehen.

2.1.3 Tier-3

Auf Tier-3 wurde basierend auf den Erfahrungen aus bwGRiD (2007 – 2012) [9, 10] die landesweit betriebene Tier-3-Rechencluster-Infrastruktur im Umsetzungskonzept bwHPC neu geordnet. Dazu werden an vier Standorten (Freiburg, Heidelberg, Tübingen, Ulm) Cluster für ausgewählte Fachdisziplinen

betrieben (bwForCluster, s. Abbildung 2) [1, 3]. Darüber hinaus wurde ein gemeinsames Cluster aller Universitäten des Landes (bwUniCluster) als Einstiegsressource für HPC und zur Grundversorgung mit Rechenleistung für alle Wissenschaftsdisziplinen etabliert, die nicht unmittelbar in den bwForClustern abgebildet sind. Unabhängig vom vorliegenden Konzept kann es spezifische lokale Rechenkapazität unterhalb von den Systemen auf Tier-3 geben, die hier nicht weiter betrachtet wird.¹

Aktuell (Stand 06/2023) stehen folgende Systeme auf Tier-3 zur Verfügung:

System	Standort	Installation	Laufzeit bis	Aufgabe
bwUniCluster2.0	Karlsruhe	2022	2027	Allgemeine Tier-3 Ebene
JUSTUS 2	Ulm	2020	2025	Theoretical Chemistry, Condensed Matter Physics, Quantum Physics
Helix	Heidelberg	2021	2026	Computational Humanities, Structural and Systems Biology, Medical Science, Soft matter
NEMO 2	Freiburg	2024	2029	Neurosciences, Particle physics, Materials science, Micro system engineering
BinAC 2	Tübingen	2023	2028	Bioinformatics, Astrophysics, Geosciences, Pharmacy, Medical Informatics

Auf Tier-3 erfolgt die fachliche Zuordnung im Detail auf Basis der DFG-Fachsystematik (vgl. Abbildung 1). Diese Zuordnung kann angepasst und erweitert werden, was thematisch im erweiterten Begleitprojekt enthalten sein wird. 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 Erfahrung 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.

¹ So gibt es beispielsweise ein Datenanalysecluster an der Hochschule Esslingen, das gebündelt Spezialbedarfe der Hochschulen für angewandte Wissenschaften adressiert und sich an den Absprachen der bwHPC-Strukturen orientiert.

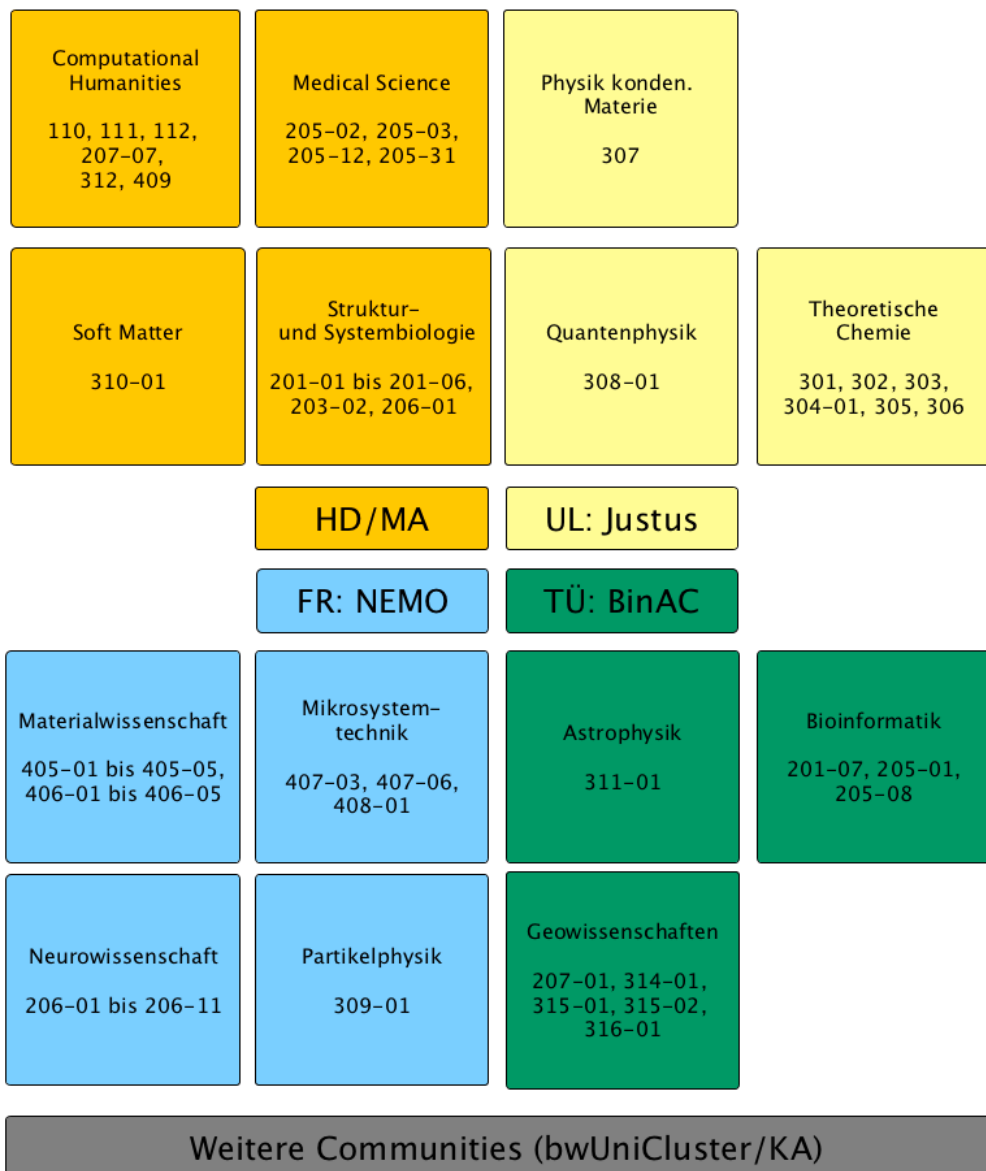


ABBILDUNG 1: FACHLICHE AUSDIFFERENZIERUNG DER HPC-SYSTEME DER EBENE 3 [1]

2.2 Datenmanagement und Speichersysteme

Ziel des Umsetzungskonzeptes des Landes für datenintensive Dienste – bwDATA [2] und des HPC-DIC-Konzepte für die Jahre 2018 – 2024 [1] war es, allen im Land vertretenen Fachwissenschaftler in herausragender Art und Weise Speicher- und Datenmanagement-Systeme für deren Bedarfe zur Verfügung zu stellen. Diese Systeme unterstützen den kompletten Scientific Data Life Cycle von der Generierung der Daten in HPC-Systemen oder Experimenten über Auswertung der Daten bis zur Ablage in Repositories und Archiv-Systemen. Um eine nahtlose Integration und eine für die Nutzenden einfache Migration von Daten zwischen allen Systemen sicherzustellen, sind die Systeme u.a. durch die Ergebnisse des Begleitprojekts bwHPC-S5 gefördert verbunden. Neben den Speichersystemen der HPC-Systeme, deren Zuordnung der fachlichen Klassifizierung der HPC-Systeme folgt (siehe voriger Abschnitt), und einem allgemeinen Archivsystem am KIT (bwDataArchive) sind die folgenden Speicher- und Datenmanagement-Systeme derzeit im Nutzerbetrieb bzw. im Aufbau:

System	Standort	Installation	Laufzeit (aktuelle HW) bis
bwSFS	Freiburg	2021	2026
bwSFS	Hohenheim	2024	2029
LSDf	Heidelberg	2019	2024
LSDf	Karlsruhe	2019	2024
bwSFS	Stuttgart	2024	2029
bwSFS	Tübingen	2021	2026

Das vorliegende Rahmenkonzept ist Teil der Strategie zum Forschungsdatenmanagement: Im Land Baden-Württemberg wird in den Jahren 2022/23 erstmals eine Forschungsdatenstrategie als Teil der Datenstrategie des Landes erarbeitet. Darin werden die strategischen Überlegungen zur Bedeutung von Forschungsdaten und zum Forschungsdatenmanagement formuliert sein.

2.3 BelWü

Mit dem Wissenschaftsnetz BelWü steht im Land eine hochperformante, hochverfügbare Infrastruktur als Basis für den Zugang zu den HPC-Systemen sowie die Vernetzung der HPC-Systeme und der Datenmanagement- und Speicher-Systeme zur Verfügung, die mit Landesförderung laufend weiterentwickelt wird. Die größtenteils exklusiv genutzte Glasfaserinfrastruktur (dark fibre) und die darauf abgestützte eigene IP-Systemplattform liefert eine große Flexibilität für die Weiterentwicklung und eine hohe Zukunftssicherheit im Hinblick auf immer höhere Bandbreitenanforderungen und neue Netzfunktionalität. Performante Übergänge zu anderen Wissenschaftsnetzen wie X-WiN und Switch sowie zu kommerziellen Netzen stellen sicher, dass international vernetzte Forschende in Baden-Württemberg optimale Voraussetzungen in der technischen Vernetzung vorfinden.

2.4 Wissenschaftsunterstützung und Governance

Komplementiert wurden diese Hardware-Strukturen durch gemeinsame, synergetische Betriebsstrukturen im Begleitprojekt bwHPC-S5 sowie eine System- und Ebenen-übergreifende Software-Versorgung. Zur besseren Absicherung der HPC-Systeme hat sich der HPC-Betrieb mit dem landesweiten Kernteam IT-Sicherheit an der Universität Heidelberg vernetzt. Alle unterstützenden Strukturen tragen erheblich zu einer Konsolidierung von Rechenressourcen und damit zu einer deutlich nachhaltigeren und effizienteren Nutzung von Mitteln und Personal bei.

Im Rahmen der bisherigen HPC-DIC-Landestrategien, der Generationen von HPC-DIC-Systemen in Baden-Württemberg und der zugehörigen Begleitprojekte (bwHPC-C5, bwHPC-S5, SDSC-BW) wurde insbesondere im Bereich der Anwendungsunterstützung ein ausgezeichnetes Umfeld auf mehreren Ebenen geschaffen. Diese sind untereinander durchlässig und ergänzen sich ohne dabei auf den genannten Bereich begrenzt zu sein. Vielmehr umfassen sie beispielsweise auch den Systembetrieb. Durch Begleitprojekte sind ebenfalls Strukturen zur technischen Weiterentwicklung der Systeme etabliert worden, in denen – etwa in Form von Technologiesprints – neue Entwicklungen im HPC-Bereich untersucht und auf ihre Einsetzbarkeit für die Systeme im Land geprüft werden.

Mit dem Modell der „Föderativen Wissenschaftsunterstützung“ existiert ein auf hohem Niveau operierender koordinierter **Nutzendensupport** für die zur Verfügung stehenden HPC- und Speicherdienste. Es gewährleistet damit eine äußerst effiziente Nutzung der Forschungsinfrastruktur und deren

kontinuierliche Optimierung auch im Sinne von „Green-IT“. Organisatorisch gliedert sich diese Unterstützungsebene in die nachfolgend genannten Aufgabengebiete bzw. Teams, an denen auch das layered-support-Modell deutlich wird.

Das **Cluster-Auswahl-Team (CAT)** unterstützt bei der Zuordnung neuer HPC- und DIC-Projekte zu den in Land vorhandenen Cluster-Standorten und gewährleistet so eine bestmögliche Arbeitsumgebung bei gleichzeitig optimaler Nutzung der Ressourcen.

bwHPC-Kompetenzzentren unterstützen mit eigenen Expertinnen und Experten in den wissenschaftlichen Bereichen, in denen besondere HPC/DIC-Anforderungen vermehrt auftreten. Derzeit sind die folgenden 7 Kompetenzzentren in der BW-Forschungs-Community integriert:

1. Bioinformatik, Pharmazie, Medizininformatik und Astrophysik
2. Computergestützte Chemie und Quantenwissenschaften
3. Geowissenschaften
4. Global Systems Science
5. Ingenieurwissenschaften
6. Elementarteilchenphysik, Neurowissenschaft, Mikrosystemtechnik und Materialwissenschaft
7. Struktur- und Systembiologie, Medical Science, Soft Matter und Computational Humanities

Tiger-Teams lösen seit Jahren erfolgreich klar umrissene Fragestellungen, in dem sie Forschende mit den notwendigen Experten im Land – etwa aus den Bereichen Hard- und Software oder auch Systembetrieb – temporär für Sprints zusammenbringen. Beispiele für Success Storys der letzten Jahre finden sich auf den bwHPC-Projektseiten [12].

Neben diesen technisch-wissenschaftlich orientierten Ebenen wird die Governance – im Sinne übergeordneter Steuergremien – wesentlich von drei Gruppen getragen.

Strategische Entscheidungen über die Fortentwicklung der Tier-3 Landes-Cluster und der landesweiten Dateninfrastruktur trifft der Arbeitskreis der Leiterinnen und Leiter der Wissenschaftlichen Rechenzentren und Informationszentren des Landes Baden-Württemberg (**ALWR-BW**), eingerichtet durch die Landesrektorenkonferenz (LRK). Anforderungen von Vorgaben anderer Steuergremien werden über den ALWR-BW und die Projektverantwortlichen koordiniert.

Für die wissenschaftliche Begleitung des Betriebs und der Weiterentwicklung definierter digitaler Forschungsinfrastrukturen ist der **Landesnutzerausschuss (LNA-BW)** zuständig. Er artikuliert die wissenschaftlichen Nutzungsanforderungen, bewertet die Wirksamkeit der Mechanismen zur Aussteuerung der Nutzung und Auslastung der Infrastrukturen, schlägt Verbesserung der Aussteuerungs- bzw. Auslastungsmechanismen vor und verfasst Empfehlungen und Stellungnahmen zu einer verbesserten wissenschaftlichen Ausrichtung der Vorhaben und Konzepte im Umfeld der HPC-DIC-Landesstrategie.

Der **Steuerkreis** digitale Forschungsinfrastruktur Baden-Württemberg repräsentiert die übergreifende Instanz sowohl der Betreiber als auch der Nutzenden der digitalen Forschungsinfrastrukturen innerhalb der HPC-DIC-Landesstrategie.

2.5 Wissenstransfer / Wirtschaft

Der Transfer von Wissen und Fähigkeiten aus der universitären Forschung in die Industrie und darüber hinaus in die Gesellschaft hat im Land Baden-Württemberg eine lange Tradition. Wurden bereits 1986 erste Schritte dazu unternommen. So war die Gründung der „Höchstleistungsrechner für Wissenschaft und Wirtschaft GmbH (HWW)“ [13] 1995 ein Meilenstein für den Wissenstransfer mit und in die Wirtschaft. Ab 2008 wurden darüber hinaus Solution Center für spezifische Problemstellungen gegründet, die Industrie und öffentliche Forschung unter einem Dach für die Forschung und Entwicklung im HPC-

DIC zusammenbringen. Heute agieren im Land Baden-Württemberg das Automotive Solution Center for Simulation (ASCS) [14], das Smart Data Solution Center (SDSC-BW) [15] und das Media Solution Center (MSC-BW) [16]. Ein Medical Solution Center ist im Aufbau begriffen. Mit dem Ziel einer speziellen Unterstützung von kleinen und mittleren Unternehmen (KMU) hat das Land Baden-Württemberg gemeinsam mit dem KIT und der Universität Stuttgart im Jahr 2011 die SICOS BW [17] gegründet.

3 Herausforderungen und Handlungsfelder

Die Landesstrategie muss Antworten auf verschiedenartige Herausforderungen, wie signifikante technologische Umbrüche, eine zunehmende Rolle der KI und einer weiteren Durchdringung der wissenschaftlichen Disziplinen finden und adressiert daher eine Reihe von Handlungsfeldern. Zur Adressierung der verschiedenen Herausforderungen bedarf es einer umfassenden Anwendungsunterstützung, ohne die die Ziele im Bereich der Energieeffizienz und Nachhaltigkeit, Nutzung neuer Hardwarearchitekturen, Verbesserung und Weiterentwicklung leistungsfähiger Forschungssoftware, effizientes Datenmanagement und Umgang mit sensiblen Daten nicht erreicht werden kann. Aspekte der digitalen Souveränität müssen mithin in jedem Handeln im Rahmen dieser Landesstrategie mitbedacht werden. Sie ziehen sich durch alle identifizierten Handlungsfelder.

3.1 Wissenschaftliche Anforderungen

Simulationsanwendungen haben in der Vergangenheit stets die verfügbare Rechenleistung ausgenutzt, und werden dies auch in Zukunft tun können, da sie immer einen Kompromiss aus der Größe des simulierten Systems und der Rechenleistung darstellen. Es gibt zudem zahlreiche konkrete Anforderungen für höhere Rechenleistung. Beispielhaft seien genannt: Forschende erwarten bei Klimasimulationen mit einer Auflösung auf der Kilometerskala eine enorme Steigerung der Vorhersagekraft, da bisher durch Parametrisierungen dargestellte Effekte durch die Simulation aufgelöst werden können. Digitale Zwillinge realer Systeme können im Allgemeinen dank höherer Rechenleistung immer realistischer werden, indem sie immer mehr und immer kleinere Teilsysteme berücksichtigen. Wissenschaftliche Experimente und Beobachtungen werden deshalb in naher Zukunft enorm wachsende Datenmengen produzieren, zudem in immer kürzerer Zeit („data avalanche“). Die Durchbrüche in ML/KI beruhen auf immer größeren Modellen, die erfolgreich trainiert werden. Die Anforderungen wachsen super-exponentiell, sowohl in Datenmenge als auch in Rechenleistung. Die digitale Transformation in der Forschung ist nicht nur auf die klassischen Communities, die HPC und DIC schon seit längerem einsetzen, beschränkt. Sie erreicht ebenso völlig neue Communities, die für den Einstieg nicht nur entsprechende Kapazitäten, sondern auch entsprechende Unterstützung benötigen. Zu nennen sind u.a. Bereiche aus den Geisteswissenschaften und den Wirtschaftswissenschaften.

Die Vervielfältigung der Forschungsbereiche geht mit der Notwendigkeit von spezifischen methodischen Ansätzen in HPC und DIC einher. Machine Learning ist ein wichtiges Beispiel in diesem Zusammenhang, da es eine enge Einbindung von Ressourcen im Bereich des Hochleistungsrechnens und des Datenmanagements erfordert. Zudem sei anzumerken, dass die enge Anbindung von Messgeräten – z.B. High Throughput Microscope – mit HPC- und DIC-Systemen auch über Standortgrenzen hinweg erforderlich ist. Die enge Verzahnung der jeweiligen Komponenten ist eine sine qua non-Bedingung, um erfolgreich diese Infrastruktur für die Wissenschaft einzusetzen.

Zuletzt sei hervorzuheben, dass wissenschaftliche Aktivitäten inzwischen äußerst interdisziplinär stattfinden, die kooperative Lösungsansätze über Communities hinweg erfordern. Dies spiegelt sich in dezentralen IT-Lösungen wider, die jedoch Kompatibilitätsvorgaben erfüllen, um entsprechenden Datenaustausch zu ermöglichen. Hierfür sind Partizipation und Mechanismen zur Durchlässigkeit einrichtungs- und länderübergreifend für die vorliegenden Forschungsbedürfnisse unbedingt zu berücksichtigen. Ziel muss es also sein, der Wissenschaft auf eine nachhaltige (sowohl im Sinne von Verlässlichkeit als auch Ressourcenschonung) Art und Weise möglichst umfangreiche und angepasste HPC/DIC-Kapazitäten (sowohl Hardware als auch Unterstützung) zur Verfügung zu stellen.

Die Weiterentwicklung der Unterstützungsstrukturen und des Supports sind innerhalb der neuen Strategie von besonderer Bedeutung, da sie die Grundlage für die effiziente Nutzung der HPC-DIC-Infrastruktur durch eine große und steigende Zahl an Forschenden und wissenschaftlichen Communities darstellt. Neben die klassische HPC-Beratung tritt zunehmend auch das Onboarding in Bezug auf die

Nutzung wissenschaftlicher IT-Infrastrukturen. Dies erfordert eine holistische Betrachtung der jeweiligen individuellen Bedarfe eines Forschungsvorhabens und im nächsten Schritt die geeignete Abbildung auf eine lokale, landesweite oder übergeordnete Forschungsinfrastruktur.

Grundsätzlich soll der bisherige Weg, Synergien durch landesweite Zusammenarbeit zu schaffen, weiter beschritten werden. Insbesondere sollen in die landesweite Nutzendenunterstützung auch weiterhin Standorte ohne HPC-DIC-Systeme einbezogen werden. Wesentliche Teile der Nutzendenunterstützung sollen auch künftig durch das Begleitprojekt bwHPC-S5 geleistet werden. Zusätzlich soll für die immer bedeutsamere Unterstützung und Begleitung des Research Software Engineering ein ergänzendes Projekt umgesetzt werden.

Um den neuen Herausforderungen und Anforderungen zu begegnen, sind konkret in den Unterstützungsstrukturen und im Begleitprojekt bwHPC-S5 folgende Weiterentwicklungen und Ergänzungen geplant:

- Weiterentwicklung der Fächerstruktur, z.B. im Hinblick auf die Exzellenzstrategie von DFG und BMBF;
- die besonderen Bedarfe der Medizin und der Wirtschaftswissenschaften;
- Weiterentwicklung der Kompetenzzentren, z.B. im Hinblick auf die Unterstützung von Methoden der KI/ML.

Eine weitere Intensivierung der Zusammenarbeit zwischen den wissenschaftlichen Anwendenden, der Methodenforschung und den Rechenzentren kann erreicht werden, indem für Wissenschaftsbereiche, die in Baden-Württemberg von besonderer Bedeutung sind, Simulation&Data Labs eingerichtet werden, in denen Forschende aus dem jeweiligen Anwendungsgebiet in engster Kooperation mit Wissenschaftlerinnen und Wissenschaftlern aus Methodenforschung und Rechenzentren gemeinsam Forschungsfragen bearbeiten. Als Blaupause können dafür die Simulation and Data Life Cycle Labs der HGF dienen, die u.a. am SCC des KIT seit vielen Jahren sehr erfolgreich arbeiten.

Insgesamt wird eine langfristig hochwertige und zuverlässig planbare Unterstützung der Forschenden jedoch nur erreichbar sein, wenn auch in den Support- und Betriebszentren eine nachhaltige Personalentwicklung stattfinden kann. Personalgewinnung ist bei einem hohen Anteil an befristeten Stellen bereits heute ein überaus ernstes Problem.

3.2 Energieeffizienz und Nachhaltigkeit

Die Sicherung der Versorgung der gesamten baden-württembergischen Wissenschaft mit jederzeit weltweit konkurrenzfähigen HPC/DIC-Systemen bringt verstärkt Herausforderungen hinsichtlich Energieeffizienz und Nachhaltigkeit mit sich. So hat sich der Energiebedarf von Hoch- und Höchstleistungsrechnern in den letzten 20 Jahren etwa verzehnfacht. Verbunden ist damit ein steigender Kühlbedarf mit allen Folgen.

Bei den Betriebskosten (insbesondere dem Energieverbrauch) der HPC/DIC-Systeme deutet sich ein Paradigmenwechsel an. Dies hängt zusammen mit dem Ende des Moore'schen Gesetzes und Dennard-Skalierung [18]. Das Moore'sche Gesetz beschrieb eine Verdopplung der Rechenleistung elektronischer Komponenten bei gleichen Kosten, Fläche und Energieverbrauch. Schon seit Längerem ist zu beobachten, dass sich Leistungssteigerungen bei Prozessoren zunehmend aus der Vergrößerung der Chip-Fläche und damit einhergehend einem steigenden Energieverbrauch speisen. Unter Annahme der vor den Preissprüngen im Jahr 2022 geltenden Strompreise entsprachen die Stromkosten für den

Betrieb eines Stücks Hardware über fünf Jahre in etwa den Anschaffungskosten. Es ist schwierig abzuschätzen, in welchem Verhältnis gegen Ende des in dieser Strategie betrachteten Zeitraums die Energiekosten zu den Anschaffungskosten stehen, aber es ist möglich, dass die Betriebskosten die Anschaffungskosten deutlich übersteigen werden.

Klimaschutz und Nachhaltigkeit erfordern es zudem, dass die Energieeffizienz zu einem wesentlichen Kriterium bei Beschaffung und Betrieb der Systeme gemacht wird. Für die Steigerung der Energieeffizienz werden der durchgehende Einsatz von hochentwickelten Kühltechniken, die Nutzung von Abwärme und die Integration der Versorgung der HPC/DIC-Systeme in die Kraft-Wärme-Kälte-Kopplung auf dem Campus angestrebt. Ziel ist es, den pro Einheit verbrauchter Energie erzeugten wissenschaftlichen Erkenntnisgewinn zu maximieren. In diesem Zusammenhang gibt die Green-IT-Landestrategie die Richtung hin zu einer Klimaneutralität klar vor. Um eine konsequent nachhaltige Aufstellung aller Betreiberstandorte zu verstärken, ist die Zertifizierung aller Zentren nach EMAS oder einem vergleichbaren Standard vorgesehen. Diese komplexe Zertifizierung soll während der Laufzeit des vorliegenden Rahmenkonzeptes in zwei Phasen realisiert werden, beginnend mit einer prototypischen Umsetzung an ausgewählten Standorten gefolgt von Ausrollen über alle Systeme.

Weiter muss, wo noch nicht geschehen, auf eine Versorgung mit Strom aus erneuerbaren Energien umgestellt werden. Außerdem müssen die Kühlkonzepte der Bestandsserverräume kritisch hinterfragt und an zeitgemäße Anforderungen angepasst werden. Auch sind verlustärmere Stromversorgungssysteme einzusetzen und, wo möglich, eine Nachnutzung von Abwärme umzusetzen. Durch die eingeübte Entkopplung von breitem Wissenschaftssupport und tatsächlichem Standort des disziplinspezifischen Gerätes kann ein zentraler Ansatzpunkt die Konsolidierung der Tier-3-Hardware an einem optimalen Betriebsstandort sein. Die Eigenschaften eines entsprechenden Standorts sind unter Berücksichtigung der Verfügbarkeit erneuerbarer Energien, effizienter Betriebsweise und weitgehender Nutzung der Abwärme zu ermitteln und schrittweise umzusetzen.

Es müssen zudem neue Betriebsmodelle untersucht und umgesetzt werden. Dazu wird eine Betrachtung von Energieverbrauch statt Rechenzeit zählen. Dies stellt einen Ansporn dar, Codes energieeffizient zu machen und z.B. auf Beschleuniger zu portieren. Erste Zentren weltweit setzen dies schon um. Prominentestes Beispiel ist Fugaku in Japan. Zudem muss das Ziel der ständigen kompletten Auslastung der Systeme hinterfragt werden. Auslegung und Betrieb der Systeme könnte sich näher an der Verfügbarkeit erneuerbarer Energien orientieren. Bei Verfügbarkeit von Überschuss-Energie (z.B. aus erneuerbaren Quellen) bzw. bei Energie-Knappheit könnte die Leistungsfähigkeit und damit der Energieverbrauch dynamisch angepasst werden.

Weitere Zielsetzungen bieten sich in der engen Nutzerunterstützung zur (energetischen) Verbesserung ihrer Workflows sowie der Auswahl der für die Forschungsfragestellung optimalen Software. Allen Maßnahmen, die Nutzende betreffen, sollten im Rahmen des Begleitprojekts Maßnahmen zur Sensibilisierung und Vorbereitung vorangestellt werden.

3.3 Systemkomponenten und -architekturen

Eine HPC/DIC-Strategie muss sich der Herausforderung stellen, flexibel auf die Entwicklung von Rechnerkomponenten und -architekturen reagieren zu können. Ein frühzeitiger Verzicht zu Gunsten neuer – und damit unerprobter – Architekturen ist ebenso problematisch wie ein zu langes Festhalten an alten Architekturen.

Mit der Entwicklung von Grafikprozessoren (Graphics Processing Units - GPU) und anderen Beschleunigerkarten für das HPC/DIC stehen Beschleunigerarchitekturen für spezifische Aufgaben zur Verfügung, die in diesem Umfeld höhere Leistung versprechen, aber auch einen höheren Programmierauf-

wand fordern. Verbessert wird die Situation dadurch, dass in Zukunft GPU-Kerne und klassische Prozessor-Kerne auf einem Chip direkt kooperieren können („Fused Parts“) und damit auch leichter zu programmieren sein werden. Sowohl Fragen der wirtschaftlichen Sinnhaftigkeit als auch der Energieeffizienz solcher Systeme stellen eine Herausforderung für Betreiber und Nutzende dar. Das Potential von Beschleunigern wird physikalisch ebenfalls limitiert.

Insbesondere unter Berücksichtigung der Energieeffizienz wird ein Eckpfeiler der Weiterentwicklung der HPC-Hardware der Einsatz neuer CPU-Architekturen und neuer Beschleuniger-Technologien sein. Die Einrichtung von innovativen Labs, die neue Technologien im Zusammenwirken zwischen den HPC-Zentren und Forschungseinrichtungen bewirkt, soll dazu führen, dass die neuesten Technologien im Hinblick auf HPC flexibel und agil erkundet werden und ggf. Teil des Angebotes der HPC/DIC-Zentren werden. So sollen beispielsweise im bwHPC-Begleitprojekt die Bereiche des Quantencomputings und der Künstlichen Intelligenz verfolgt werden, speziell auch mit Blick auf mögliche neue Betriebs- und Einsatzmöglichkeiten. Die Erprobung sollte Hardware, Algorithmen, Software und Anwendungen umfassen.

Hinsichtlich neuartiger Computerkonzepte ist an erster Stelle das Quantum Computing zu nennen, obwohl aktuell unklar ist, ob, wann und wofür Quantencomputer für wissenschaftliche Zwecke produktiv nutzbar sein werden. Neue, potenziell disruptive Technologien sollen kontinuierlich beobachtet und erprobt werden. Sollte sich insbesondere zeigen, dass in Zukunft betriebsfähige Quantenrechner verfügbar sind, muss das Landeskonzept deren Integration, auch unterstützt durch das Begleitprojekt, ermöglichen.

Neben Quantencomputern sind dies vor allem Tensor Processing Units oder monolithische Chips. In diesem sich schnell entwickelnden Feld ist jedoch nicht absehbar, welche Technologien sich durchsetzen werden.

In zunehmendem Maß stellen kommerzielle Anbieter Ressourcen in der Cloud zur Verfügung. In dem Ausmaß wie Provider auch Lösungen für rechenintensive Anwendungen ins Portfolio übernehmen, stellt sich die Frage danach, welche Bereiche des HPC/DIC von öffentlichen Providern wie Rechenzentren abgedeckt werden sollten und welche Bereiche an private Anbieter übergeben werden können. Es sind zudem Mischformen denkbar, z.B. „Cloud Bursting“, d.h. bei hoher Auslastung werden insbesondere kleinere Capacity Jobs in ein Cloud-Rechenzentrum geschoben. Neben wirtschaftlichen Fragen sind bei diesen Überlegungen Datenschutz und -hoheit zu berücksichtigen.

Teil der Weiterentwicklung der Systeme wird die verstärkte Bereitstellung interaktiver, cloud-artiger Nutzungsmöglichkeiten sein. Dies kann unter anderem durch die Definition von dafür spezifisch definierten Partitionen auf dem HPC-System erfolgen. Neben Anpassungen des Scheduling wird auch eine Öffnung für weitere Betriebssysteme, z.B. Windows, erfolgen. Somit können auch weitere Communities, z.B. im Bereich der Wirtschaftswissenschaften, erschlossen werden. Diese Flexibilität soll unter anderem durch die Verwendung von Virtualisierung und Container-Lösungen erreicht werden. Diese Technologien sollen auch Möglichkeiten des sogenannten insitu-computing eröffnen, womit Datentransfers gezielt optimiert werden können.

3.4 Software

Software ist zu einer Schlüsselkomponente der wissenschaftlichen Arbeit geworden, und es gibt heute kaum noch eine Forschungsdisziplin, die nicht den Einsatz von Software erfordert. In Software sind Algorithmen und Wissen kodiert. Während quelloffene Software eine direkte Mitgestaltung und Weiterentwicklung durch die Wissenschafts-Community erlaubt, bestehen bei herstelleregebundener Software Risiken, die aus den sich regelmäßig ändernden Preis- und Lizenzmodellen erwachsen. Die Herausforderung besteht darin, neue Maßstäbe für die Entwicklung herausragender Forschungssoftware

zu setzen und das Research Software Engineering in eine neue Ära zu führen. Besondere aktuelle Bedeutung bekommt dies auch vor dem Ziel der Steigerung der Energieeffizienz und der Nachhaltigkeit insbesondere der HPC-Systeme. Soll neben „mehr HPC-Cycles pro kWh“ auch „mehr Erkenntnisgewinn pro kWh“ erzielt werden, ist Energie-optimierte Forschungssoftware ein entscheidender Baustein.

Großgeräte, Dienste und weitere Ressourcen im Zusammenhang mit Forschungssoftware werden von der wissenschaftlichen Gemeinschaft allgegenwärtig genutzt. Daher muss Forschungssoftware die gleichen strengen Anforderungen erfüllen, die Forschende an ihre Daten, Proben, Geräte und Infrastrukturen stellen. Software muss – wie jede andere Infrastruktur auch – kontinuierlich entwickelt, gewartet und unterstützt werden, manchmal über Jahrzehnte hinweg. Erfolgreiche und nachhaltige Softwareprojekte beruhen oft auf starken, florierenden und meist internationalen Gemeinschaften und erfordern stets eine langfristige Finanzierung. So fordert beispielsweise das niederländische eScience Centre in seinem Strategiepapier: „Forschungssoftware muss auf politischer Ebene und in der Praxis gleichberechtigt mit Forschungsdaten und Publikationen behandelt werden“ [19]. Auch das britische Software Sustainability Institute stellt schlicht fest: „Better Software, Better Research“ [20].

Auch auf Landesebene ist Forschungssoftware als lebenswichtige Infrastruktur für die Bedürfnisse der modernen Wissenschaft nachhaltig zu behandeln, um damit wissenschaftliche Erkenntnisse und Durchbrüche zu sichern. Exzellente Forschungssoftware kann sich auch zu erfolgreichen Dienstleistungen entwickeln. Insbesondere quelloffene Software bietet vielfältige Ansatzpunkte zur Steigerung der Recheneffizienz und die permanente Anpassung auf neue Hardwarearchitekturen und Betriebsmodelle. Die (Weiter-)Entwicklung offener Forschungssoftware und Algorithmen vertieft das Verständnis von Forschenden und schafft eine breite Grundlage für den Know-How-Transfer in Wirtschaft und Gesellschaft. Damit können weitere Impulse für alle unsere Forschungsdisziplinen gesetzt werden und damit die führende Rolle des Landes Baden-Württemberg im wissenschaftlichen Rechnen in Deutschland und darüber hinaus erhalten und ausgebaut werden.

Die Herausforderungen bestehen in:

- der Schaffung eines Bewusstseins dafür, dass Forschungssoftware ein wesentlicher Bestandteil der Landes-Forschungsinfrastruktur ist, der entsprechend implementiert und betrieben werden muss;
- der Wertschätzung von Forschungssoftware als legitimer Forschungsoutput;
- einem nachhaltigen und professionellen Ansatz für die Entwicklung und das Management von Energie-optimierter Forschungssoftware, etabliert durch geeignete Unterstützungsstrukturen und -dienste sowie mit motivierten und entsprechend bezahlten Forschungssoftware-Ingenieuren;
- der Unterstützung von Communities bei der Erstellung optimierter Software, insbesondere bezogen auf den Energiebedarf, für ihre Forschungsarbeit und bei Angeboten dazu bei allen Partnern und Interessenten; Unterstützung ist hier insbesondere erforderlich für hochskalierbare parallele Software, Adaption auf moderne Rechnerarchitektur (z.B. GPU), optimierte numerische Methoden für Simulation, Software Sustainability and Performance Engineering, Continuous Integration Services und Continuous Development.

Die bisherigen Erfahrungen im HPC-Bereich haben deutlich gemacht, dass die Berücksichtigung der zugrundeliegenden Hardware alleine zum erfolgreichen Einsatz in der Forschung nicht reicht. Die Software-Entwicklung spielt eine entscheidende Rolle. Das gilt nicht nur zur Performance-Gewinnung, sondern auch bzgl. der Energieeffizienz. Die Software-Entwicklung ist insofern herausfordernd, als die zugrundeliegenden Rechner-Architekturen von HPC-Systemen zunehmend heterogen sind, was zu einer sehr hohen Komplexität führt. Dies spiegelt sich insbesondere in der Verwendung von Beschleunigern

wie GPUs und FPGA, die abgestimmt mit den CPUs arbeiten müssen, wider. Die Mehrkern-Technologie, die inzwischen allgegenwärtig in Rechner-Architekturen Einzug gefunden hat, erfordert spezifische Konzepte zur optimalen Nutzung. Inzwischen ist allgemein bekannt, dass das größte Potenzial zur Leistungsgewinnung tatsächlich in der Softwareentwicklung steckt. Durch die erhöhte Komplexität der heutigen Rechner-Architekturen kann man nicht davon ausgehen, dass ein Mehr an Leistung ohne Zutun der Forschenden automatisch gewonnen werden kann. Nicht selten sind erhebliche Anpassungen der Software hierfür erforderlich. Softwareentwicklung ist ein zunehmend wichtiger Bestandteil der Forschungsaktivitäten, aber wenn sie nicht nachhaltig durchgeführt wird, ist das Ergebnis oft nicht wartbare Software und nicht reproduzierbare Wissenschaft. Dies ist auf einen Mangel an Software-Engineering-Ausbildung für Forschende, begrenzte Mittel für die Weiterentwicklung und Wartung bestehender Software und wenige feste Softwareentwicklerstellen zurückzuführen.

Die anvisierte Strategie zielt darauf, die Entwicklung wissenschaftlicher Software zu fördern und dadurch zu verbessern, um reproduzierbare Wissenschaft und Software-Nachhaltigkeit zu gewährleisten. Die geplanten Maßnahmen sollen als Bindeglied zwischen den verschiedenen wissenschaftlichen Disziplinen dienen und so die Zusammenarbeit und interdisziplinäre Forschung ermöglichen.

Zur Software-Strategie soll ein drei-Säulen-Konzept umgesetzt werden, das folgende Schwerpunkte beinhaltet: Entwicklung, Lehre und Öffentlichkeitsarbeit.

Laut dem UK Research Software Survey 2014 [21] geben 69 % der Forscher an, dass ihre Forschung nicht mehr praktikabel wäre, wenn sie keine Forschungssoftware mehr nutzen könnten. 56 % der 417 befragten Forscher entwickeln ihre eigene Forschungssoftware. Das vorliegende Konzept sieht entsprechend vor, dass die Software-Entwicklung für wissenschaftliche Forschungsaktivitäten für HPC-Systeme sehr stark gefördert und unterstützt werden soll. Dies gilt sowohl für die Entwicklung von Open Source numerischen Buildingblocks, die an die jeweiligen Systeme angepasst werden, als auch für die Anpassung von existierender Software auf die vorhandene Rechner-Architektur. Diese Aktivitäten werden nicht nur von den HPC-Zentren getragen, sondern sollen die Kompetenzen der jeweiligen Universitäten im Bereich des wissenschaftlichen Rechnens und der Software-Entwicklung heranziehen.

Weiterhin sollen Kurse zu allen Aspekten der nachhaltigen Softwareentwicklung sowie Mentoring-Programme insbesondere für Pre- und Postdocs der Universitäten des Landes Baden-Württemberg angeboten werden. Dies gilt gleichermaßen für erfahrene Forschende, die entsprechende weiterführende Schulungen und Konferenzen im Bereich der Software-Entwicklung wahrnehmen können sollen.

In diesem Sinne ist ein landesweites Format zu initiieren, mit der Zielsetzung, Initiativen im Bereich der Software-Entwicklung und der Lehre zu fördern. Dabei soll ein entsprechendes HPC-Software-Ökosystem mit hoher Sichtbarkeit durch konkrete Projekte unter dem Dach des vorliegenden Rahmenkonzeptes landesweit etabliert werden. Diese sollen den Schwerpunkt auf langfristige Softwarelösungen legen. Über diese Maßnahmen soll weiterhin eine Kompetenz ausgebaut werden, die von allen Forschenden auf den HPC-Systemen des Landes Baden-Württemberg im Sinne einer Beratung bzw. einer Unterstützung bei der Software-Entwicklung in Anspruch genommen werden kann. Schwerpunktmäßig sollen hierbei zunächst die Bereiche KI, Machine Learning und Data Analytics herangezogen werden.

3.5 Datenmanagement und Speichersysteme

Es ist zu erwarten, dass die in zukünftigen Messungen, Beobachtungen oder Simulationen produzierten Datenmengen exponentiell wachsen werden. Ein wesentlicher Energieverbrauch tritt jedoch erst dann auf, wenn Daten prozessiert bzw. analysiert werden. Hingegen erfordern die Fach-Community orientierte Spezialisierung der Compute- und Weiterverarbeitungs- und Datenanalyse-systeme sowie

das Forschungsdatenmanagement eine zunehmend bessere Verzahnung der digitalen Forschungsinfrastrukturen über den gesamten Datenlebenszyklus. Forschende erheben oder erhalten Daten aus Quellen, die ebenso wie Weiterverarbeitung, Visualisierung und anschließende Bereitstellung für Dritte nicht mehr am eigenen Standort lokalisiert sein müssen. Mit dem zunehmenden Ausbau der landesweiten Speichersysteme nimmt deren Nutzung zu. Hierbei sind die Vorstellungen der Forschenden aus leicht integrierbaren Workflows mit direktem Datenzugriff mit den Möglichkeiten verteilter Systeme abzugleichen. Das vorliegende Rahmenkonzept ist in diesem Zusammenhang Teil des künftigen Rahmenkonzeptes zur Forschungsdateninfrastruktur (siehe 2.2).

Den Herausforderungen im Bereich der Forschungsdaten wird durch eine konsequente Weiterentwicklung der Datenmanagement- und Speichersysteme in der BW-Datenföderation begegnet. Während hier die Grundlagen durch eine einheitliche Authentifizierungs- und Autorisierungsstruktur gelegt sind, erfordern viele Workflows das Kopieren der Daten auf den jeweils optimierten Speicherbereich beispielsweise eines HPC-Systems oder zur späteren sicheren Ablage in einem Datenpublikations-Repository. Eine solche "Copy"-Datenföderation sollte für geeignete Usecases um eine "Mount"-Datenföderation erweitert werden, die das Einspielen der Daten an einer Stelle und die direkte Nutzung an anderer Stelle erlaubt. Das zügige Bewegen großer Datenmengen erfordert sowohl die Verfügbarkeit schneller Datennetze als auch die Integration in zukünftige Scheduling- und Staging-Systeme. Ein weiterer Schritt besteht im verstärkten Aufbau einer Object Storage Föderation, welches einerseits den Datenfernzugriff im Vergleich zu Netzwerkdateisystemen erheblich vereinfacht, jedoch ein Umdenken auf Anwendungs- und Workflow-Seite erfordert. Repositorien für Datenpublikation und Langzeitarchivierung müssen mittel- und langfristig das FDM durch einerseits generische als auch auf die Anforderungen der Fach-Communities abgestimmte Metadatenbereitstellung über intelligente Suchfunktionen und wohldefinierte APIs unterstützen.

3.6 Umgang mit sensiblen Daten etwa aus der Medizin

Die Forschenden in den bwHPC-Domänen Bioinformatik, Neurowissenschaften, Medical Science aber auch Computational Humanities arbeiten regelmäßig mit sensiblen Datensätzen. Dies bezeichnet unter anderem alle personenbezogenen Daten, die Rückschlüsse auf die Herkunft zulassen und umfasst genetische Daten, biometrische Daten und Gesundheitsdaten. Auf anderen Wissenschaftsfeldern kommen alle sozio-ökonomische Daten mit Bezug zu den religiösen, politischen oder weltanschaulichen Überzeugungen einer Person hinzu.

Die Verarbeitung von sensiblen Daten unterliegt der EU-Verordnung 2016/679 und der Datenschutzgrundverordnung (DSGVO), die das Vorliegen eines Zwecks, wie auch eine Einwilligung der betroffenen Personen vorschreiben.

Im Rahmen der Fortschreibung der HPC/DIC-Strategie sind Konzepte notwendig, die die Speicherung und Verarbeitung derartiger personenbezogener Daten auch an anderen Betreiberstandorten innerhalb der Gruppe der Landesuniversitäten ermöglicht.

Im Kontext von bwHPC ist der Zweck zumeist durch das Forschungsinteresse gegeben, wobei nicht für jeden Datensatz eine Einwilligung für jeden Forschungszweck (general consent) vorliegt. Für jeden Datensatz sollte auch ein eindeutiges Votum der zuständigen Ethikkommission und des entsprechenden Data-Access-Committees vorliegen. Diese Aspekte liegen aber vollständig in der Verantwortung der Forschenden und können durch die Betreiber der bwHPC-Ressourcen nicht abgedeckt oder durchgesetzt werden. Im Sinne der DSGVO sind die Betreiber für gewöhnlich Auftragsverarbeiter der Forschenden, alleine schon dadurch, dass sensible personenbezogene Daten gespeichert werden. Dies bedingt,

dass zwischen der Einrichtung der Forschenden und der des Betreibers ein Vertrag zur Datenverarbeitung im Auftrag geschlossen werden muss. Gehören beide zur selben Einrichtung, sollten zumindest die wesentlichen Abläufe dokumentiert werden (Standard Operating Procedures, SOP).

Die letzten Jahre haben gezeigt, dass auf Seite der Forschenden, wie auch der Betreiber Unsicherheiten bestehen, was die rechtlichen Grundlagen und ihre praktische Umsetzung angeht. An einigen Standorten, gerade jene mit starkem biomedizinischen Fokus, wurde dieser Themenkomplex jedoch angegangen. Die Erfahrungen hieraus definieren die zukünftigen Aktivitäten in bwHPC.

Die wichtigste Aktivität ist der Stärkung der Expertise bei allen Beteiligten. Dies macht unmittelbare persönliche Interaktion notwendig, da jedes Projekt mit sensiblen Daten seine eigenen Charakteristika aufweist. Dieser individuelle Support entspricht wenigstens dem Niveau eines TigerTeams und ist entsprechend personalintensiv.

Auf Seiten der Betreiber besteht vor allen Dingen die Notwendigkeit, die technischen und organisatorischen Maßnahmen in der Form von projektbezogenen SOPs zu dokumentieren. Hierin werden Ansprechpartner, Zuständigkeiten und Verantwortlichkeiten definiert und konkrete Maßnahmen zur Absicherung der (virtuellen) Umgebungen festgelegt. Je nach Komplexität des Forschungsvorhabens und Schutzbedarf des sensiblen Datensatzes variiert die Detailtiefe dieser SOP-Dokumente, wobei durch Standardisierung Synergieeffekte zu erwarten sind.

Auf technischer Ebene sind die derzeit in Betrieb befindlichen bwForCluster nur bedingt zur Prozessierung sensibler Daten mit hohem Schutzbedarf geeignet. Beim üblichen Einsatz von verteilten Clusterfilesystemen ist keine hinreichende Separation der Daten und der Ausführungsumgebung möglich. Für zukünftige Systeme ist hierfür schon vor der Beschaffung der Hardware darauf zu achten, dass Datenseparationsstrategien wie Encryption-at-Rest, feingranulare ACLs, VLAN-Separation, und ähnliches, nicht nur möglich sind, sondern auch mit dem Betriebskonzept der Cluster harmoniert.

Entsprechend der Verpflichtung, die die Hochschulen für ihre Organisationseinheiten eingegangen sind, führen alle Cluster-Betreiber ein Informationssicherheitsmanagementsystem (ISMS) ein, um vor allen Dingen die organisatorischen Aspekte von zeitgemäßer Datensicherheit umsetzen zu können. Die Zertifizierung nach ISO27001 oder BSI-C5 wird nachdrücklich allen Standorten empfohlen, nicht zuletzt, um nachprüfbar zu dokumentieren, dass geeignete Maßnahmen zur Datensicherheit gelebt werden. Dies wird das Vertrauen der Forschenden in die bwHPC-Infrastruktur und die Kompetenz der Betreiber nachhaltig stärken.

3.7 Digitale Souveränität

Digitale Souveränität bedeutet die Fähigkeiten und Möglichkeiten, Aufgaben in einer digitalen Welt selbständig, selbstbestimmt und sicher erledigen zu können. Für Universitäten und Forschungseinrichtungen ist das die Grundlage von freier Forschung, womit Transparenz, Handlungsfähigkeit und uneingeschränkte Wahl der Werkzeuge zur Leitmaxime digitaler Souveränität und damit der Landesstrategie HPC-DIC werden. In einer globalisierten Welt liegen in der Regel nicht alle notwendigen Bausteine in einer Hand, so dass zwangsläufig ein Abwägungsprozess zwischen verfügbaren digitalen Bausteinen (Chips, Netzwerk, Software), Zulieferern (geopolitische Erwägungen) und Formen der Bereitstellung (HPC on-premises, cloud) entsteht. Die Offenheit wissenschaftlicher Ergebnisse und speziell von Codes (Open Science, Open Source) stellt besonders im wissenschaftlichen Bereich ein wichtiges Kriterium für den Abwägungsprozess dar, etwa wenn es um Fragen der Weiterentwicklung oder Nachvollziehbarkeit von Forschungsdaten geht. Zur Förderung der digitalen Souveränität sind eine gemeinsame Strategieentwicklung mit Partnerinstitutionen für ein abgestimmtes Handeln und universitätsübergreifende Kooperationen wichtig, um Synergien zu erzielen und gemeinsame Standards umzusetzen. In diesem Kontext können regelmäßige Folgenabschätzungen vorgenommen, der notwendige und akzeptierte

Grad der Unabhängigkeit verhandelt, die Wirtschaftlichkeit und Nutzerfreundlichkeit von IT-Lösungen beurteilt werden. Schließlich sind der Kompetenzerhalt und die Kompetenzentwicklung der Verantwortlichen, Betreibenden und Forschenden zentral zur Ausübung souveräner Entscheidungen.

Die breite Ausbildung des wissenschaftlichen Nachwuchses und die Förderung von offener Wissenschaft und Open Source schafft notwendige Voraussetzungen für die digitale Souveränität in den Bereichen HPC/DIC. Die öffentlich geförderte Wissenschaft kann hier eine Vorreiterrolle einnehmen, die Wege zur Umsetzung der verschiedenen Aspekte digitaler Souveränität aufzeigt und damit ihre Vorbildfunktion für eine Anwendung in Wirtschaft und Zivilgesellschaft ausfüllt.

3.8 Strukturen und Organisation

HPC/DIC im Land Baden-Württemberg ist in internationale, europäische und deutsche Strukturen eingebunden. Auf deutscher Ebene sind das weiterhin das Gauss Centre for Supercomputing (GCS) sowie das Nationale Hochleistungsrechnen (NHR). Auf europäischer Ebene entwickelt sich dagegen mit dem Joint Undertaking EuroHPC [22] eine neue, zusätzliche Struktur. Aus der Sicht der Nutzenden ergeben sich damit vier organisatorische Ebenen. Die lokale Ebene der Universität, die Landesebene BW (Tier-3 bwHPC), die Bundesebene (Tier-1 GCS und Tier-2 NHR) sowie die europäische Ebene (EuroHPC). Die Herausforderung für den Hochschulstandort BW und seine Betreibereinrichtungen besteht darin, die durchgängige Nutzung dieser Ebenen für die Wissenschaft im Land verfügbar zu machen. Es muss ein Ziel sein, diese Ebenen möglichst mitzugestalten. Für EuroHPC, GCS und NHR ist das aktuell der Fall.

Die Fortführung der bisherigen Konzepte im Land zu HPC und DIC steht durch diese Entwicklung vor der Herausforderung einer weiteren Integration mit neuen Strukturen wie NHR und der Etablierung neuer und vertiefter Schnittstellen zu speziellen Communities, wie dies etwa durch de.NBI für die Bioinformatik bereits erfolgreich umgesetzt ist.

Mit der Nationalen Forschungsdateninfrastruktur (NFDI) entstehen bundesweite Verbünde einzelner Fach-Communities zur übergreifenden Standardisierung und Weiterentwicklung des Forschungsdatenmanagements. Viele HPC-Communities werden in Zukunft NFDI-Dienste beziehen und zu diesen beitragen, so wie etliche HPC- und DIC-Provider in Baden-Württemberg für die NFDI Dienstleistungen erbringen. Dieses erfordert einerseits eine Abstimmung der Strategien, um Doppelentwicklungen zu vermeiden, und bietet andererseits die Chance, die etablierten Dienste des DIC bundesweit anzubieten. Dieses erfordert nicht nur eine Integration in übergreifende AAI-Strukturen, sondern ebenfalls gesicherte Betriebsmodelle, das Angebot der Verarbeitung sensibler Daten und die Möglichkeit der Abrechnung dieser Angebote.

Die Universitäten und Hochschulen des Landes Baden-Württemberg legen im Jahr 2023 durch eine Rahmenvereinbarung die Basis für eine Formalisierung ihrer IT-Kooperationen vor, welche auch die rechtlichen Rahmenbedingungen abdeckt. Innerhalb dieser IT-Allianz für die Wissenschaft können IT-Dienste angeboten und von anderen genutzt werden. Rechtsgrundlage hierfür ist die im Landeshochschulgesetz vorgesehene Kooperation nach LHG §6 Absatz 1. Über die IT-Allianz werden auch Fragen wie die Auftragsverarbeitung nach Artikel 28 DSGVO und Haftungsfragen grundlegend geklärt.

Die erfolgreiche bisherige Kooperation für das Hochleistungsrechnen soll zentraler Bestandteil der IT-Allianz werden, die damit der bewährten wissenschaftlichen Kooperation den notwendigen formalen und rechtlichen Rahmen gewährt.

Wissenschaftlerinnen und Wissenschaftler an den Universitäten und Hochschulen des Landes haben durch die IT-Allianz den formalen Nutzungsrahmen für die Dienstleistungen der Forschungsinfrastrukturen aus dem vorgelegten Konzept, ebenso regelt die Vorgabe der IT-Allianz das Kostenmodell und gibt Rechtssicherheit in diversen Fragen.

Kooperationen über die Landesgrenzen hinaus – etwa im Bereich der NFDI – werden durch die IT-Allianz nicht abgedeckt. Da sich diese auf die öffentlich-rechtlichen Hochschulen des Landes Baden-Württemberg bezieht. Hier sind ergänzende Vereinbarungen notwendig, wie sie aktuell beispielsweise durch die Nationale Forschungsinfrastruktur oder auch im Rahmen der Science Data Center des Landes entwickelt werden.

4 Quellen und Referenzen

- [1] Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS²DM), ALWR-BW, 2018, <http://dx.doi.org/10.15496/publikation-27872>
- [2] Rahmenkonzept der Hochschulen des Landes Baden-Württemberg für datenintensive Dienste – bwDATA (2020-2024), ALWR-BW, 2020, <http://dx.doi.org/10.15496/publikation-55923>
- [3] Hartenstein, H., Walter, T. und Castellaz, P. (2013). Aktuelle Umsetzungskonzepte der Universitäten des Landes Baden-Württemberg für Hochleistungsrechnen und datenintensive Dienste. PIK-Praxis der Informationsverarbeitung und Kommunikation 36(2): 99-108
- [4] Konzeptpapier des LNA-BW zur HPC-DIC-Landesstrategie 2025 – 2032 vom 27.09.2022
- [5] https://eurohpc-ju.europa.eu/index_en
- [6] <http://www.gauss-centre.eu>
- [7] <http://www.nhr-verein.de/>
- [8] <https://www.nfdi.de>
- [9] <https://www.degruyter.com/document/doi/10.1515/pik-2013-0007/pdf>
- [10] <https://publikationen.bibliothek.kit.edu/1000039516/3083829>
- [11] http://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher/
- [12] https://www.bwhpc.de/success_stories.php
- [13] <https://www.hww.de/>
- [14] <https://www.asc-s.de/>
- [15] <https://www.sdsc-bw.de/>
- [16] <https://www.msc-bw.com/>
- [17] <https://www.sicos-bw.de/>
- [18] Shalf J. 2020 The future of computing beyond Moore’s Law. Phil. Trans. R. Soc. A 378: 20190061. <http://dx.doi.org/10.1098/rsta.2019.0061>
- [19] <https://doi.org/10.5281/zenodo.3378572>
- [20] <https://www.software.ac.uk/resources/publications/better-software-better-research>
- [21] S. J. Hettrick et al, <https://zenodo.org/record/1183562>
- [22] <https://eurohpc-ju.europa.eu/>