

Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS²DM)

Vorgelegt von den Leiterinnen und Leitern der Rechenzentren bzw. Informationszentren der Universitäten des Landes Baden-Württemberg (ALWR-BW) in Abstimmung mit dem Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg

Zur Begutachtung durch die Deutsche Forschungsgemeinschaft (DFG)

Revision 1.0

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

Prof. Dr. Gerhard Schneider, Universität Freiburg

Prof. Dr. Vincent Heuveline, Universität Heidelberg

Karl-Wilhelm Horstmann, Universität Hohenheim

Prof. Dr. Bernhard Neumair, Karlsruher Institut für Technologie

Petra Hätscher, Universität Konstanz

Dr. Josef Kolbitsch, Universität Mannheim

Dr. Simone Rehm, Universität Stuttgart

Prof. Dr.-Ing. Dr. h.c. mult. Michael Resch, Universität Stuttgart

Prof. Dr. Thomas Walter, Universität Tübingen

Prof. Dr.-Ing. Dr. h.c. Stefan Wesner, Universität Ulm

Peter Castellaz, Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg
(ständiger Gast)

Zusammenfassung

Computational Sciences¹ 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“² 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 Wissenschaft³ 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 sinnvoll von einzelnen Universitäten oder Forschungsinstitutionen 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 auf Basis ihrer HPC- und DATA-Konzepte verschränken. Kooperative Lösungen helfen die beschriebenen Herausforderungen zu bewältigen und versprechen einen institutionen- und disziplinübergreifenden Mehrwert.

Für die Periode von 2018 bis 2024 ist es das Ziel aller beteiligten Akteure, den beschrittenen Weg der Kooperation gemäß der HPC Landesstrategie⁴ 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)⁵ mitzuwirken. Im Sinne eines integrierten Ansatzes werden die bestehenden Konzepte für HPC, DIC und LS²DM weiterentwickelt und in einer gemeinsamen Strategie zusammengeführt. Gleichzeitig werden die Grundlagen für eine frühe Beteiligung am Aufbau einer NFDI geschaffen und erforderliche Infrastrukturen bereitgestellt.

¹ <https://www.wissenschaftsrat.de/download/archiv/4032-14.pdf>

² <http://www.wissenschaftsrat.de/download/archiv/4488-15.pdf>

³ Tony Hey, *The Fourth Paradigm: Data-Intensive Scientific Discovery* (2009)

⁴ <https://mwk.baden-wuerttemberg.de/de/service/presse/pressemitteilung/pid/hpc-landesstrategie-eine-halbe-milliarde-euro-fuer-digitale-infrastruktur-und-supercomputer/>

⁵ vgl. hierzu auch RfII, <http://www.rfii.de/?wpdmdl=2269>

Die wesentlichen Ziele der geplanten Weiterentwicklung sind

- Festigung der für die Forschung und Wissenschaft in Baden-Württemberg bereitgestellten Dienste und Infrastrukturen in den Bereichen HPC, DIC und LS²DM auf dem etablierten hohen Niveau und aktive Unterstützung bei Entwicklung und Standardisierung einer NFDI
- Einbettung der Landesstrategie HPC, DIC und LS²DM in den bundesweiten und europäischen Kontext
- Bündelung und Integration der individuellen Aktivitäten in diesem Themenbereich aus den einzelnen wissenschaftlichen Communities
- Anpassung der Kooperations-, Governance- und Support-Strukturen
- Ausbau und Schärfung der fachlichen Ausdifferenzierung für qualitativ hochwertige und fundierte Nutzerunterstützung

Das hier vorgelegte Konzept konkretisiert gleichzeitig das Rahmenkonzept bwDATA⁶ 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.

Zusammengefasst ergeben sich folgende Maßnahmen:

- Erweiterung der Aufgabenbereiche der Kompetenzzentren auf Datenhaltung und Einrichtung weiterer Kompetenzzentren
- Bedarfsgerechter Ausbau und Erneuerung der HPC-Systeme aller drei Ebenen und Datenmanagementsysteme
- Entwicklung einer landesweiten Datenföderation
- Fortführung des Begleitprojektes mit angepassten Schwerpunkten und Ausstattung im Hinblick auf die neuen Anforderungen einer Integration von HPC, DIC und LS²DM (Abschnitt siehe V.5)

In den folgenden Abschnitten werden nach einer einleitenden Motivation die aktuelle Ausgangssituation und insbesondere die Erfahrungen mit bereits etablierten Diensten beschrieben. Davon ausgehend werden die Weiterentwicklung des Konzepts und die geplanten künftigen Dienste, Infrastruktur und Governance motiviert und dargelegt.

⁶ Rahmenkonzept der Hochschulen des Landes Baden-Württemberg für datenintensive Dienste – bwDATA (2015-2019): <http://dx.doi.org/10.15496/publikation-21187>

I. Einleitung

Die Bedeutung von Computational Sciences und HPC-Systemen als deren Fundament nimmt nach wie vor zu. In seinen Empfehlungen zur „Finanzierung des Nationalen Hoch- und Höchstleistungsrechnens in Deutschland“⁷ betont der Wissenschaftsrat erneut die Unverzichtbarkeit des Hoch- und Höchstleistungsrechnens für die Wettbewerbsfähigkeit Deutschlands in zahlreichen Wissenschaftsfeldern und empfiehlt die grundsätzliche Beibehaltung der HPC-Leistungspyramide. Von besonderer Bedeutung ist dabei, dass alle Ebenen der Leistungspyramide aufeinander abgestimmt und miteinander koordiniert sind (s. Abbildung 1).

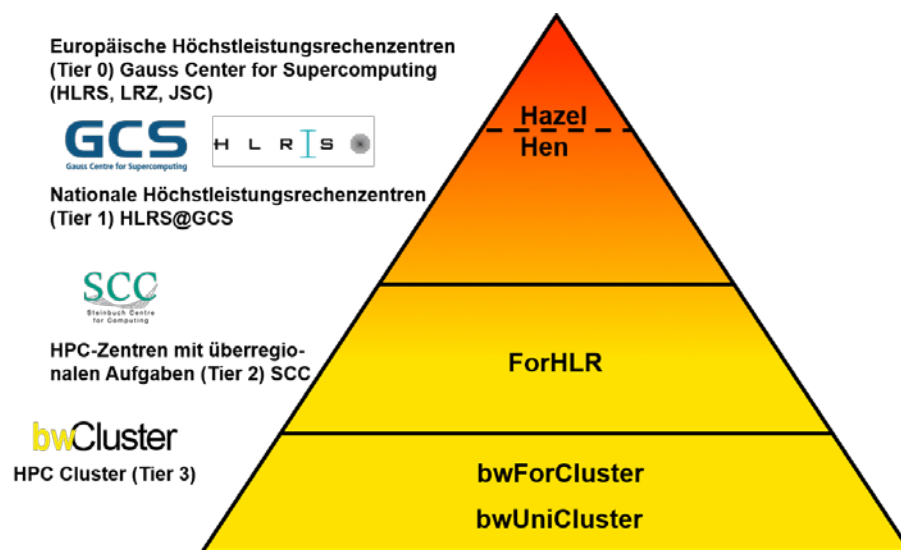


ABBILDUNG 1: LEISTUNGSEBENEN (TIER) DES HOCHLEISTUNGSRECHNENS IN BADEN-WÜRTTEMBERG IM RAHMEN DER DEUTSCHEN LEISTUNGSPYRAMIDE DES HPC.

Gleichzeitig nimmt in allen Bereichen der Wissenschaft das Volumen an digitalen Daten ständig zu. Die Daten werden in verschiedenen Formaten erzeugt, gesammelt und aufgezeichnet. Die Gewinnung neuer wissenschaftlicher Erkenntnisse durch Auswertung dieser Daten und durch Verknüpfung von großen Datenmengen aus unterschiedlichen Quellen wird heute nach Theorie, Experiment und Simulation als 4. Säule der Wissenschaft⁸ bezeichnet. Der zunehmende Einsatz von Data Intensive Computing wird auch vom Wissenschaftsrat als aktuelle Entwicklung besonders hervorgehoben.

⁷ <http://www.wissenschaftsrat.de/download/archiv/4488-15.pdf>

⁸ Tony Hey, The Fourth Paradigm: Data-Intensive Scientific Discovery (2009)

II. Ausgangssituation in Baden-Württemberg

II.1 HPC-Leistungspyramide in Baden-Württemberg

Gemäß der Wissenschaftsrats-Empfehlung wird der höchsten Leistungsklasse im HPC (Tier-1) vor allem der Bereich Capability Computing zugeordnet, während die Systeme der nächstniedrigeren 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 Communities.

Baden-Württemberg hat mit seinem bwHPC-Konzept⁹ einen von der DFG als beispielgebend bezeichneten Rahmen¹⁰ für die Kooperation und Abstimmung zwischen den genannten Ebenen und damit ein in Deutschland herausragendes Umfeld für die Wissenschaften geschaffen. Auch auf Bitte der DFG wurden die Umsetzungskonzepte bwHPC und bwDATA als Best Practice Beispiele veröffentlicht.¹¹ Die genannte Leistungspyramide wird gemäß dem bwHPC-Konzept wie folgt umgesetzt:

- Auf Tier-0 (als einer der drei Partner des GCS) und Tier-1 übernimmt das Höchstleistungsrechenzentrum Stuttgart (HLRS) der Universität Stuttgart die Rolle eines nationalen und europäischen Höchstleistungsrechenzentrums.
- Auf Tier-2 übernimmt das Steinbuch Center for Computing (SCC) am Karlsruher Institut für Technologie (KIT) die Rolle eines Hochleistungsrechenzentrums und Zentrums für datenintensives Rechnen.
- Auf Tier-3 wurde basierend auf den Erfahrungen aus bwGRiD die landesweit betriebene Tier-3-Rechencluster-Infrastruktur im Umsetzungskonzept bwHPC neu geordnet. Dazu werden an vier Standorten (Freiburg, Heidelberg/Mannheim, Tübingen, Ulm) Cluster für ausgewählte Fachdisziplinen betrieben (bwForCluster, s. Abbildung 2). Darüber hinaus wurde ein gemeinsames Cluster aller Universitäten des Landes (bwUniCluster) und Einstiegsressource für HPC und zur Grundversorgung mit Rechenleistung für alle Wissenschaftsdisziplinen etabliert, die nicht unmittelbar in den bwForClustern abgebildet sind.

Somit wurden in Baden-Württemberg Infrastrukturen geschaffen, die es erlauben, jegliche wissenschaftliche Fragestellung mit Bezug zu Computational Sciences in voller Tiefe wie auch Breite zu bearbeiten.

⁹ Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für das Hochleistungsrechnen: <http://dx.doi.org/10.15496/publikation-21185>

¹⁰ Schreiben der DFG vom 21.03.2013

¹¹ Siehe auch: Hartenstein, H., Walter, T., Castellaz, P., „Aktuelle Umsetzungskonzepte der Universitäten des Landes Baden-Württemberg für Hochleistungsrechnen und datenintensive Dienste“, PIK - Praxis der Informationsverarbeitung und Kommunikation, Band 36, Heft 2, Seiten 99-108, DOI: 10.1515/pik-2013-0007

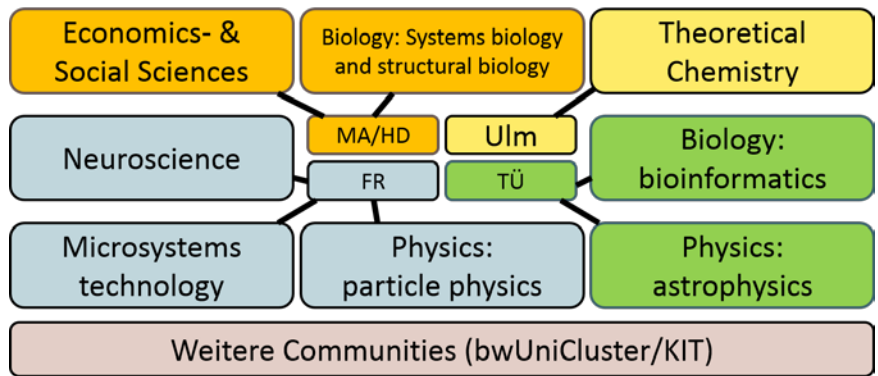


ABBILDUNG 2: FACHLICHE AUSDIFFERENZIERUNG DER HPC-SYSTEME DER EBENE 3 (BISHERIGER STAND)

II.2 Data Intensive Computing und Large Scale Scientific Data Management

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).

Abbildung 3 zeigt beispielhaft, wie in Forschungsprozessen Daten generiert, verarbeitet und transferiert werden und welche IT-Systeme, insbesondere Datenspeichersysteme, beteiligt sind:

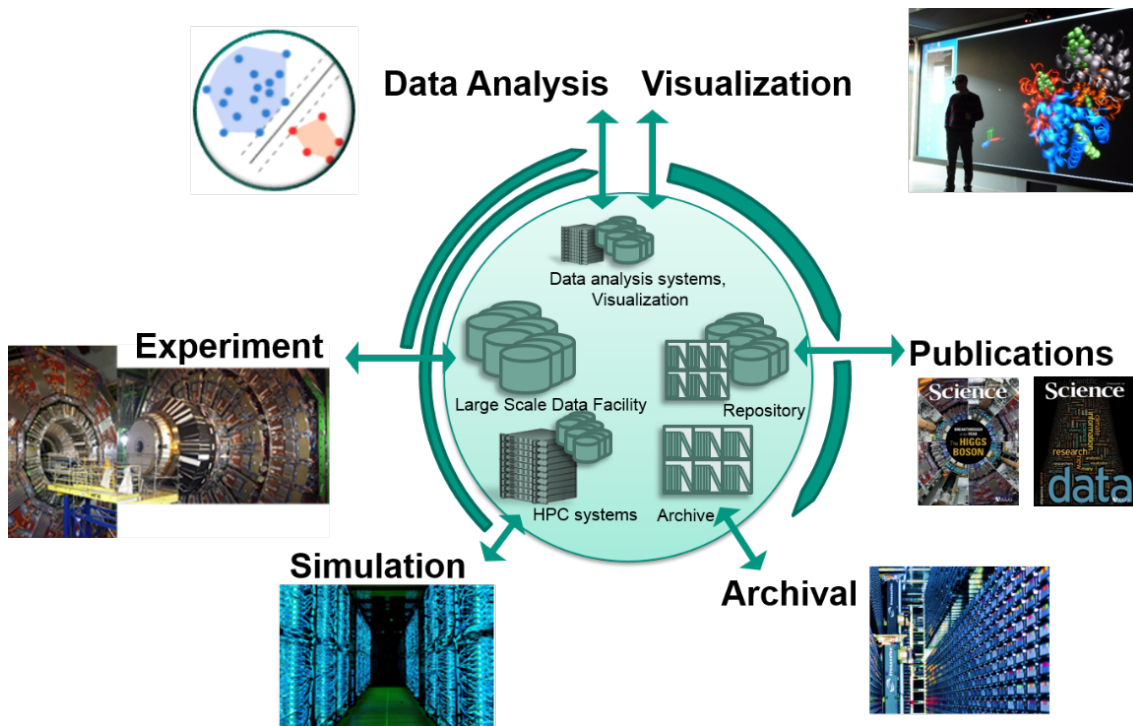


ABBILDUNG 3: DATA LIFE CYCLE VON WISSENSCHAFTLICHER FORSCHUNG.

- Simulationen auf HPC-Systemen liefern 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 des Forschungsvorhabens werden die gesammelten Daten für eine spätere erneute Nutzung archiviert und dazu an ein für langfristige Archivierungszwecke ausgelegtes Datenspeichersystem übertragen.

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 Communities sind. Zudem generiert die Interaktion der Wissenschaftler mit den wissenschaftlichen Daten durch Qualitätssicherung und Anreicherung der Metadaten echten Mehrwert.

In Baden-Württemberg wurde für datenintensives Rechnen und das Management großvolumiger wissenschaftlicher Daten bereits eine föderierte Infrastruktur etabliert:

- **Data Intensive Computing:** Das Tier-2-HPC-System unterstützt durch seine technische Ausprägung insbesondere Data Intensive Computing. Es bietet ein integriertes Datenmanagement mit sehr großen Hintergrundspeichern und durchgängiger Unterstützung des gesamten Scientific Data Life Cycle. Es besteht eine nahtlose Verknüpfung des Systems mit Smart Data Solution Center Baden-Württemberg (SDSC-BW) und der LSDF in Karlsruhe/Heidelberg.
- **Datenmanagementsysteme:** Die von KIT und der Universität Heidelberg betriebene LSDF unterstützt datenintensive Experimente z.B. in den Lebenswissenschaften. Große Mengen an Daten aus leistungsfähigen Messgeräten wie Gen-Sequenzierern können direkt in der LSDF im Rahmen des Dienstes *SDS@hd* am Standort Heidelberg und *LSDF Online Storage* am Standort Karlsruhe abgelegt, auf HPC-Systemen weiterverarbeitet und am Arbeitsplatz visualisiert werden, wobei die Datensicherheit über Zugriffsregelungen erhalten bleibt. Diese Art des komplexen Datenmanagements ist für die Nutzer damit vollständig transparent und leistet einen weiteren wichtigen Beitrag zur Unterstützung des Scientific Data Life Cycle.
- **Spezialsysteme für Datenanalyse:** Am SCC werden im Rahmen der Forschungsplattform Smart Data Innovation Lab (SDIL) mehrere leistungsstarke Datenanalysesysteme

für die Forschung betrieben, die von Bund, Land Baden-Württemberg und IT-Unternehmen gefördert werden. Neben performanter Hardware zeichnet sich die Forschungsplattform durch die Verfügbarkeit von hochentwickelten Software-Systemen wie die SAP HANA In-Memory-Computing-Plattform, die Software AG Terracotta In-Memory Data Fabric Plattform, die IBM InfoSphere BigInsights Analytik-Plattform, IBM SPSS Modeler und Server sowie Huawei FusionInsight. Mit Hilfe dieser Plattform unterstützen SCC und die SICOS-BW GmbH im Rahmen des SDSC-BW gemeinsam kleine und mittelständische Unternehmen bei der Hebung von Wertschöpfungspotentialen mittels innovativer Smart Data Technologien.

- **Integration von HPC und Data Analytics:** Die 2016 erfolgte Beschaffung zweier Rechensysteme zur Datenanalyse, die in den bestehenden Höchstleistungsrechner des HLRS integriert sind, wird flankiert durch ein vom Land Baden-Württemberg, dem HLRS, der Firma Cray und weiteren Industriepartnern getragenes Projekt zur Integration von High Performance Data Analytics (HPDA) und HPC. Dadurch ist am Standort Stuttgart auf der Ebene von Tier-0 und Tier-1 die Grundlage für eine weitere Konvergenz von Computing und Data Analyse entstanden.
- **Langfristige Datenaufbewahrung:** Wichtiges Element der Unterstützung des Scientific Data Life Cycle ist die digitale Archivierung auch großvolumiger wissenschaftlicher Daten. Neben den im Land zur Verfügung stehenden disziplinspezifischen Archivierungsmöglichkeiten wurde von KIT und HLRS mit bwDataArchiv ein generischer Langzeitdatenspeicher für alle Wissenschaftsdisziplinen der baden-württembergischen Universitäten realisiert. Ein weiteres Landesprojekt (bwDataDiss) ermöglicht den Wissenschaftlerinnen und Wissenschaftlern die zu Dissertationen gehörenden Forschungsdaten zu speichern, zu beschreiben und öffentlich zugänglich zu machen. Damit wird eine Öffnung in Richtung Data Repositories wie z.B. RADAR geschaffen.

Forschungsdatenmanagement

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*¹² die Bedeutung des Forschungsdatenmanagements für eine nachhaltige Nutzung von Forschungsdaten betont und konsequent zwei Förderprogramme (Forschungsdatenmanagement und Virtuelle Forschungsumgebungen) eingerichtet.¹³ Das Forschungsdatenmanagement (FDM), die nachhaltige und zukunftsorientierte Organisation von Forschungsdaten, ihre Verfügbarmachung und Publikation sind heute ein Fokus moderner Forschungsprozesse. Die Werkzeuge des Forschungsdatenmanagements sind insbesondere notwendig, wenn die Vorgaben individueller Datenmanagementpläne (DMP) umgesetzt werden müssen und dazu ein Metadatenmanagement erfordern.

¹² siehe *E-Science Wissenschaft unter neuen Rahmenbedingungen – Fachkonzept zur Weiterentwicklung der wissenschaftlichen Infrastruktur in Baden-Württemberg, Stuttgart 2014*

¹³ siehe insbesondere <https://www.forschungsdaten.info>

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 Forschungsdatenmanagement verschränken. So sollen im Rahmen des Gesamtkonzepts 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 Forschungsdatenmanagement integriert anbieten.

II.3 Begleitprojekt bwHPC-C5 Coordinated Compute Cluster Competence Centers

Aus der Erkenntnis heraus, dass signifikante Investitionen in Forschungsinfrastrukturen alleine, d.h. ohne entsprechende Unterstützung und Steuerung, nicht die gewünschte Wirkung entfalten, wurde zeitgleich mit bwHPC das Begleitprojekt bwHPC-C5 etabliert. Diese "Coordinated Compute Cluster Competence Centers"¹⁴ bestehen aus Kompetenzzentren mit vom Land geförderten und von den Universitäten komplementär finanzierten Personalstellen, die – nach fachlichen Schwerpunkten aufgeteilt – Hilfestellung bei Nutzung und Anpassungen sowie Unterstützung der Forschenden und Nachwuchswissenschaftler leisten.

Das neu zu beantragende Begleitprojekt *bwHPC-S5 Scientific Simulation und Storage Support Service* soll die notwendigen technischen Komponenten und Entwicklungen bereitstellen, aufbauend auf bwHPC-C5 und erweitert um die Aspekte Data Intensive Computing und Large Scale Scientific Data Management.

III. Bisherige Erfahrungen und Schlussfolgerungen

III.1 Allgemeine Erfahrungen

Bisher sind primär die Aktivitäten am HLRS in Stuttgart und am SCC in Karlsruhe in nationale und internationale Strukturen eingebunden. Die Digitalisierungsaktivitäten des Bundes und der europäischen Kommission machen es aber zunehmend notwendig, auch auf Ebene 3 eine überregionale Sichtweise einzunehmen. Erste Schritte, etwa mit der Teilnahme von Freiburg, Heidelberg und Tübingen an de.NBI,¹⁵ wurden bereits eingeleitet.

Die bisher erfolgte Implementierung der oben genannten Infrastruktur basierte auf zwei getrennten Konzepten: bwHPC für das Hochleistungsrechnen und bwDATA für große wissenschaftliche Daten. Aufgrund der großen Bedeutung von Daten in vielen Bereichen der Wissen-

¹⁴ <http://www.bwhpc-c5.de>

¹⁵ <https://www.denbi.de>

schaft ist diese Trennung nicht mehr zielführend, worauf auch die DFG in ihren Stellungnahmen zu bwHPC und bwDATA hingewiesen hat. Es muss daher eine noch engere Kopplung der Rechen- und Speichersysteme realisiert werden, und die angebotenen Dienstleistungen nicht als getrennte Einheiten, sondern als Teile eines typischen Workflows der Wissenschaftler angesehen werden, die möglichst nahtlos integriert sein müssen.

Die fachliche Ausdifferenzierung von Supportstrukturen wie Kompetenzzentren und Tiger-Teams hat sich bewährt. Diese im Bereich HPC bewährten Strukturen müssen angemessen auf die Bereiche DIC und LS²DM erweitert werden. Dies erfordert auch eine entsprechende Anpassung der etablierten Governance-Strukturen und der Zusammenführung bislang getrennter Steuer- bzw. Lenkungsstrukturen.

Auch wenn das Thema Forschungsdatenmanagement in seiner ganzen Breite nicht Teil dieses Umsetzungskonzeptes ist, müssen die Dienste und Infrastrukturen für die Bereiche HPC, DIC und LS²DM die notwendigen Voraussetzungen für ein FDM sehr großer Daten schaffen. Ausgehend von bestehenden föderierten Infrastrukturen wie LSDF und bwDataArchiv und den Planungen für weitere bwSFS (Storage for Science) werden die Dienste für die Unterstützung von FDM-Funktionen ausgerichtet sein müssen.

III.2 Erfahrungen spezifisch für HPC Tier-3

Die im bwHPC Konzept vorgesehene fachliche Zuordnung von Rechenvorhaben zum passenden Forschungscluster konnte für mehr als 2/3 der Rechenvorhaben innerhalb von 3 Tagen und für 90% innerhalb von 10 Tagen umgesetzt werden. Auch wenn hier gegenüber dem direkten Zugang zum bwUniCluster eine minimale Verzögerung entsteht, konnte die Wartezeit bis zum Zugang gegenüber bwGrid schon allein durch den Verzicht auf den zertifikatsbasierten Zugang erheblich reduziert werden. Für einige wenige Vorhaben war eine fachliche Zuordnung aufgrund einer interdisziplinären Ausrichtung schwieriger, oder ein Nutzer benötigte abweichend von der üblichen Betriebsumgebung seines Fachs z.B. eine abweichende Hardwareausstattung.

Um dies in Zukunft noch besser zu adressieren und auch im Berichtswesen eine klare Aufgliederung zu erreichen, werden die unterschiedlich granularen Fachzuordnungen der Cluster in eine auf der DFG Fachsystematik¹⁶ aufbauenden Gliederung überführt. Für Entwicklungs- oder Evaluierungszwecke oder bei speziellen Anforderungen an Hardware und Umgebung kann zum Nutzen der Anwender von der Fachsystematik eines bwForClusters abgewichen werden. Konkret bedeutet dies, dass für Methoden- und Softwareentwicklung (z.B. Fachbereich Mathematik und Informatik) die Möglichkeit des Zugangs auf mehreren Rechensystemen nicht nur möglich, sondern im Sinne einer breiten Nutzbarkeit der Methoden sogar explizit als wünschenswert angesehen wird. In gleichem Maße kann beispielsweise für die Bewertung von neuen Softwareversionen von VASP mit GPGPU Unterstützung einem Chemiker mit Zugang in

¹⁶ http://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher

Ulm auch ein Zugang in Tübingen ermöglicht werden. Nach gemeinsamer Bewertung in Form eines Tiger Teams kann dies auch für die Produktion ermöglicht werden und bei der zukünftigen Systemarchitektur von Nachfolgesystemen berücksichtigt werden.

Zur Bewertung der Auslastung und Nutzung wurde für alle Rechensysteme unter Berücksichtigung der konkreten Architekturen und Betriebsmodelle ein einheitliches Verfahren zur Erhebung der Daten im Projekte bwHPC-C5 realisiert. Diese Daten stehen über ein Web-Portal Betreibern und Steuerkreisen zur Verfügung, werden aber auch zusätzlich für die Betreiber, den bwHPC Steuerkreis und den Landesnutzerausschuss regelmäßig aufbereitet und kommentiert (z.B. Wartungsfenster oder besonderen Ereignisse). Auf Basis dieser Daten konnten unterstützt durch den Landesnutzerausschuss (LNA-BW) auch neue Fachgruppen für bwForCluster identifiziert werden, die aufgrund ihrer starken Nutzung des bwUniClusters in einem bwForCluster aufgenommen werden sollen. In gleichem Maße können mit diesen Daten auch Nutzer identifiziert werden, die von Tier-2 oder Tier-0/1 Ressourcen profitieren können.

Im Jahr 2017 wurde eine durch den Lehrstuhl für Soziologie der Universität Hohenheim wissenschaftlich unterstützte Umfrage unter den bwHPC-Nutzern durchgeführt. Ziel war es, Daten zur Zufriedenheit mit den Diensten, zu den Infrastrukturen und zur Benutzerunterstützung zu erheben. Mehr als 80 % der Nutzer haben die bereitgestellte Dienstleistung als gut oder sehr gut beurteilt. Die Wartezeit der Jobs im Batch-System wurde parallel mit der Bewertung 2,4 auf einer Notenskala von 1 (sehr gut) bis 5 (mangelhaft) bewertet. Die Ergebnisse der Umfrage bildeten sowohl die Basis für die weitere Planung, als auch für die Priorisierung der Verbesserungen.

III.3 Erfahrungen spezifisch für Large Scale Scientific Data Management

Im Gemeinschaftsprojekt „Large Scale Data Facility - Phase 2 (LSDF2)“ wurde durch die Universität Heidelberg und das Karlsruher Institut für Technologie (KIT) eine Speicherlösung für große wissenschaftliche Daten an beiden Standorten aufgebaut. Sie wird seit Anfang 2017 als Landesdienst betrieben. Auch die Etablierung einer zentralen Infrastruktur für die Langzeit-speicherung (bwDataArchiv) hat sich aus technischer Sicht als zweckmäßig erwiesen und kann bereits heute die Anforderungen vieler Wissenschaftsdisziplinen erfüllen.

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.

Um diese Anforderungen besser und systematischer zu erfassen wurde analog zu einem Rechenvorhaben auch das Konzept eines Speichervorhabens eingeführt. Für die Verwaltung der

Heidelberger LSDF₂ für den Speicherdienst „SDS@hd - Scientific Data Storage“¹⁷ wurde das Konzept Anfang 2017 erfolgreich umgesetzt. Durch diesen Dienst erfolgt Zugangsmanagement, Nutzerverwaltung, Berichtswesen und die Verwaltung von Zugriffen für die Ressource LSDF₂.

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 Daten-Föderation müssen daher auch die Möglichkeiten geschaffen werden, lokale Datenquellen performant anzubinden und beim Datentransfer Mechanismen wie Anonymisierung/Pseudonymisierung oder Aggregation zu unterstützen.

Mit den bisherigen Maßnahmen ist eine leistungsfähige Infrastruktur entstanden, die Forschern neben den lokalen Datenspeichern an den Rechensystemen die Möglichkeiten bietet, mittelfristig oder auch langfristig Daten geeignet abzulegen. In einem weiteren Schritt sind nun darauf aufbauend die Dienste von Datenspeicherung hin zu Datenmanagement zu entwickeln und geeignet in Prozesse zu integrieren, die nicht nur die Wissenschaftler unterstützen, die annotierten Daten in einem langfristig nutzbaren Format abzulegen, sondern auch deren Nachnutzung ermöglichen. Für einen Regelbetrieb und für die Unterstützung des FAIR-Prinzips (Findable, Accessible, Interoperable, and Reusable)¹⁸ muss eine Integration sowohl in lokale als auch in fachspezifische Repositorien im Bereich des Forschungsdatenmanagements etabliert werden.

III.4 Erfahrungen spezifisch für Datenanalyse

Im Rahmen des Betriebs der Forschungsplattform SDIL und des SDSC-BW konnten am SCC umfangreiche Erfahrungen mit Cloud-artigen Bereitstellungsmodellen für Analysesysteme und deren Integration in die vorhandene HPC- und Dateninfrastruktur gewonnen werden. Derartige Modelle werden künftig sowohl für die Datenanalyse auf dafür spezialisierten, leistungsfähigen Plattformen als auch für die Bereitstellung von Rechenkapazität als Abrundung des HPC-Angebots an Bedeutung gewinnen.

Für die Auswertung der weiter zunehmenden Datenmengen und -vielfalt sind die bisher etablierten Rechensysteme in ihrer Konzeptionsphase nur eingeschränkt ausgelegt worden und haben sich mehr am etablierten Bedarf für Batch-Umgebungen für Simulationen orientiert. Ob die speziellen Anforderungen durch angepasste Software-Stacks oder in Kombination mit spezieller Hardware abgebildet werden können, oder ob eine Co-Location von Daten und Analysesystemen notwendig ist, muss weiter anhand von verschiedenen Anwendungsszenarien untersucht werden. Neben den einschlägigen Erfahrungen der SDIL und am HLRS, sowie in der

¹⁷ <https://www.urz.uni-heidelberg.de/de/sds-hd>

¹⁸ z.B. <https://www.nature.com/articles/sdata201618>

bwCloud¹⁹ müssen integrierte Softwarestacks weiter untersucht werden, um hier die Anforderungen noch besser an die Bedürfnisse der bestehenden und neuer Nutzergruppen anzupassen.

IV. Weiterentwicklung

Für die kommende Periode von 2018 bis 2024 soll die beschriebene, bisher sehr erfolgreiche Strategie weiterentwickelt werden, um die Dienste und die Infrastruktur im Bereich HPC, DIC und LS²DM für die Wissenschaft in Baden-Württemberg weiter auszubauen und stärker als bisher zu integrieren.

Exemplarische Darstellung der Herausforderung getrennter Lebensdauern und Zugriffsregelungen eines Rechenvorhabens und der aus dem Rechenvorhaben entstehenden Daten („Speichervorhaben“)

Die bisher erreichte Integration erlaubt beispielsweise, dass ein Nutzer der Universität Konstanz jeweils mit seinem lokalen Universitätsaccount auf den bwForCluster BinAC in Tübingen, die LSDF SDS@hd in Heidelberg und auch dem bwDataArchiv in Karlsruhe Rechenvorhaben bzw. Speichervorhaben registrieren kann. Darüber hinaus hat der Nutzer ebenfalls mit dem lokalen Universitäts Account in der bwCloud in der Zone Freiburg virtuelle Server für eine Scientific Workflow Software reserviert. Verlässt der Nutzer die Universität, wird automatisch der lokale Account stillgelegt, eine Nutzung der HPC- und Cloud Ressourcen ist damit nicht mehr möglich. Dies gilt momentan in gleicher Form für die gespeicherten Daten.

Hier sind Erweiterungen in Prozessen und auch in der technischen Infrastruktur zu realisieren die z.B. auch nach Ablauf des lokalen Accounts und damit des HPC-System-Zugangs die Nachnutzung und Verfügbarkeit der Daten möglich macht. Dazu gehören unter anderem die Möglichkeit, Daten zwischen verschiedenen Speichersystem zu verschieben, zeitlich gestaffelte Zugriffsregeln für Daten während des Data Life Cycle und die Aufbereitung der Daten für eine längerfristige Archivierung.

IV.1 Ziele

Ausgehend von den bisher gemachten Erfahrungen beim Betrieb der Dienste im Bereich der Simulation und Daten soll die bisherige vorwiegend technische Integration auf Basis der föderierten Identitäten 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.

¹⁹ <http://www.bw-cloud.org>

Schematische Darstellung eines künftigen Data-Life-Cycles am Beispiel eines Workflows der Bioinformatik

Der Data-Life-Cycle beginnt mit der Datenerzeugung durch einen Sequencer.

Aus einer Gewebeprobe werden mit Hilfe eines Sequencers die genetischen Informationen extrahiert und nach einer initialen Verarbeitung als Daten auf dem Primärsystem gespeichert. Da es sich um hoch sensible Patientendaten handelt, muss bereits auf dem Primärsystem eine Pseudonymisierung der Daten stattfinden.

Als Vorbereitung zur Datenanalyse werden mehrere Patientendatensätze als Kohorte aggregiert. Damit ist die Datenerzeugung für die Kohorte abgeschlossen, die aggregierten Datensätze werden jetzt – automatisch oder manuell initiiert – auf das bwSFS Speichersystem übertragen.

Zu Beginn der Datenanalyse werden die aggregierten Datensätze automatisiert auf das parallele Dateisystem des bwForClusters BinAC übertragen. Dort findet die bioinformatische Aufarbeitung und Analyse der Datensätze statt. Die dabei erzeugten sekundären Daten werden von mehreren Wissenschaftlern geprüft, weiter analysiert und wissenschaftlich gemäß den Vorgaben des Datenmanagementplans (DMP) annotiert.

Anschließend findet für die sekundären Daten (Ergebnisse) und Änderungen an bestehenden Daten eine Rückführung auf das bwSFS-Speichersystem statt. Erfordert die wissenschaftliche Analyse eine weitere Aufarbeitung der sekundären Daten wird direkt ein weiterer Analysezyklus initiiert.

Nach Abschluss der Datenanalyse werden alle Primär- und Sekundär-Daten in einem sogenannten Sicherungsspace gesichert und mittels persistenter Identifikatoren für die Datenpublikation aufbereitet. Die Datenarchivierung zur langfristigen Sicherung und Nachnutzung der Forschungsdaten erfolgt je nach den Vorgaben des DMP in einem Repositorium oder einem Archiv.

Mit der Nachnutzung der Forschungsdaten im gleichen oder anderen Projekten beginnt anschließend ein neuer Data-Life-Cycle.

Es ergeben sich daher für die Themenfelder HPC, DIC und LS²DM folgende konkrete Ziele:

- Ausbau der bestehenden landesweiten Support-Strukturen (Kompetenzzentren, Tiger Teams und Ticket-System), um neue Themenbereiche und diensteübergreifende Unterstützungsprojekte zu ermöglichen. Dazu wird das das Nachfolgeprojekt bwHPC-S5 gegenüber bwHPC-C5 thematisch erweitert: Es umfasst zukünftig alle Themenfelder der wissenschaftlichen Rechen- und Datendienste und leistet Support in all diesen Belangen.
- Neben der Einbindung der Dienste auf Ebene 3 in den bundesweiten und europäischen Kontext im Bereich HPC und die gegebene Durchlässigkeit zu höheren Leistungsstufen mit den Zentren in Karlsruhe (SCC) und Stuttgart (HLRS) müssen auch Entwicklungen wie eine künftige Nationale Forschungsdateninfrastruktur (NFDI) und die globale Vernetzung der Wissenschaft, wie bereits mit de.NBI exemplarisch realisiert, noch stärker berücksichtigt werden.

- Stärker abgestimmter Betrieb von Rechen- und Dateninfrastruktur der Ebene 3 auch bei der zeitlichen Planung von Ausbau der Kapazität und Dienstangeboten.
- Anpassung der etablierten Governance auf die erweiterten Inhaltsbereiche.
- Kontinuierliche Validierung und Anpassung der fachlichen Ausdifferenzierung der Infrastruktur und Unterstützungsmaßnahmen, insbesondere bei neuen Themenschwerpunkten oder Nutzergruppen.

V. Maßnahmen

V.1 Erneuerung/Entwicklung der HPC-Systeme der Ebenen 1 und 2

Am HLRS wird eine Weiterführung des baden-württembergischen Engagements für Tier-0 und Tier-1 in GCS und über GCS auch im europäischen Kontext von PRACE verfolgt. Um die internationale Konkurrenzfähigkeit in der Forschung sicherzustellen, muss ein zu beschaffendes System ebenfalls weltweit konkurrenzfähig sein. Gleichzeitig muss durch Abstimmung mit Partnern in Deutschland und in Europa eine optimale Fokussierung der verfügbaren Ressourcen erfolgen. Das HLRS wird 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. Der mit dem aktuellen Tier-2-System eingeschlagene methodische Weg – Tier-2-HPC mit Fokus auf datenintensivem Rechnen und hoher Energieeffizienz – soll durch das Nachfolgesystem konsequent fortgeführt und weiterentwickelt werden. Das SCC wird die Beschaffung von modernsten Hochleistungsrechnern und Datenmanagementsystemen an diesen Zielen ausrichten und den Rahmen des nationalen Hoch- und Höchstleistungsrechnens (NHR) berücksichtigen. 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. Den Wissenschaftsrats-Empfehlungen sowie der bisherigen Praxis folgend wird das Tier-2-System auch künftig für Forschende außerhalb Baden-Württembergs verfügbar sein.

V.2 Erneuerung/Weiterentwicklung 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 Anforderungen der jeweiligen Wissenschafts-

Communities in Bezug auf Hardware-Auslegung und die Software-Ausstattung werden dokumentiert in die Planungen aufgenommen. Die Erneuerung der Systeme in Ulm, Heidelberg/Mannheim, Freiburg und Tübingen wird im Rahmen des Programms „Forschungsgroßgeräte“ der DFG nach Art. 91b GG beantragt. 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, vgl. Kapitel V) wird dazu fortgeführt.²⁰ Parallel dazu sind im Zeitraum bis 2024 zwei weitere Ausbaustufen des allgemeinen bwUniClusters am KIT im Rahmen des Programms „Großgeräte der Länder“ nach Art. 143c GG vorgesehen.

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.

Parallel zur fachlichen Ausrichtung bieten die Systeme auf Tier-3 Ressourcen für die Methodenforschung an. Dabei sollen u. a. die Weiterentwicklung und Optimierung von hochgradig skalierbaren numerischen Lösern auf Hochleistungsrechnern unterstützt werden. Methodisch sollen die Forschungsaktivitäten rund um das Datenmanagement für HPC-Anwendungen ausgebaut werden. Diese Forschungsaktivitäten, die in Bezug auf die fachliche Ausrichtung einen Querschnittscharakter aufweisen, wurden bereits erfolgreich mit dem IWR²¹ am bwForCluster in Heidelberg/Mannheim durchgeführt.

²⁰ vgl. hierzu Wiebelt Bernd, Janczyk Michael, von Suchodoletz Dirk, Aertsen Ad, Rotter Stefan, Quast Günter, Schumacher Markus, Greiner Andreas, „Strukturvorschlag für eine bwHPC-Governance der ENM-Community“ bzw. Wesner Stefan, Walter Thomas, Wiebelt Bernd, von Suchodoletz Dirk, Schneider Gerhard, „Strukturen und Gremien einer bwHPC-Governance Momentaufnahmen und Perspektiven“ In: Kooperation von Rechenzentren Governance- und Steuerung - Organisation, Rechtsgrundlagen, Politik 2016, de Gruyter,

²¹ <https://typo.iwr.uni-heidelberg.de/home/>

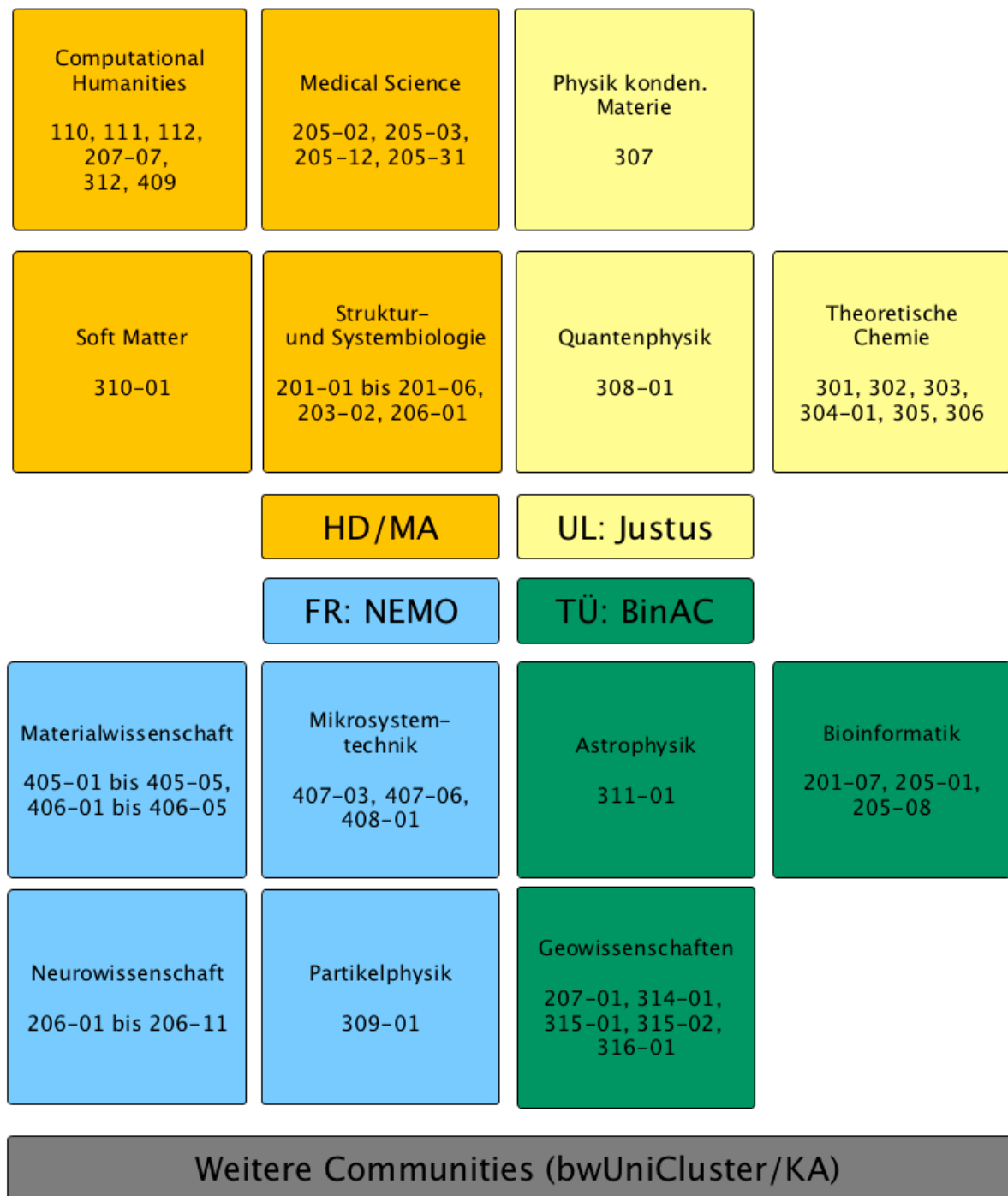


ABBILDUNG 4: KÜNFTIGEN FACHLICHE AUSSDIFFERENZIERUNG DER HPC-SYSTEME DER EBENE 3

Auf Tier-3 soll die fachliche Zuordnung künftig im Detail auf Basis der DFG-Fachsystematik²² erfolgen (vgl. Abbildung 4). Diese Zuordnung kann angepasst und erweitert werden, was thematisch im erweiterten Begleitprojekt enthalten sein wird. Insbesondere bei Erweiterung und

²² http://www.dfg.de/dfg_profil/gremien/fachkollegien/faecher/

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 hinaus gehender 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 Ersatzbeschaffungen und Hardwareerneuerungen hierfür werden daher im Rahmen des Programms „Großgeräte der Länder“ nach Art. 143c GG mit Eigenbeteiligung der jeweiligen Hochschule in zwei Stufen erfolgen. Bei der Auslegung des Systems wird geprüft, ob für die beteiligten Hochschulen jeweils parallele Jobs oder Durchsatz im Vordergrund stehen, und das Kommunikationsnetzwerk wird entsprechend ausgelegt.

Die im Rahmen des Landesprojektes bwCloud aufgebaute Infrastruktur 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.

V.3 Erneuerung und Weiterentwicklung von datenintensivem Rechnen und dem Management großer Mengen wissenschaftlicher Daten

In Abschnitt II.2 wurde beispielhaft dargestellt, wie in Forschungsprozessen Daten generiert, verarbeitet und transferiert werden und wie IT-Systeme und insbesondere Datenspeichersysteme dabei involviert sind. In Baden-Württemberg sind zur Unterstützung dieser Prozesse bereits Lösungen etabliert. Um aber die Unterstützung des datenintensiven Rechnens und des Managements großer Mengen wissenschaftlicher Daten im Interesse der Wissenschaft weiter voranzutreiben, zu verbessern und in die Breite zu tragen, sind drei Handlungsstränge erforderlich:

- Bedarfsgerechte Erneuerung vorhandener Datenmanagement- und Datenspeichersysteme sowie Aufbau zusätzlicher Systeme
- Verbindung dieser Systeme zur BaWü-Datenföderation
- Weiterentwicklung der Supportstrukturen und Integration in den Scientific Data Life Cycle

Die Forschungsplattform SDIL, die mehrere leistungsstarke Datenanalysesysteme u. a. für SDSC-BW bereitstellt, soll erneuert und weiterentwickelt werden:

- Die auf der Plattform verfügbare Datenanalysesoftware wird erneuert und erweitert.
- Die Plattformhardware wird bedarfsgerecht erneuert.
- Die Betriebsmodelle, die sich erheblich von den seit langem bekannten und ausgereiften Betriebsmodellen aus dem HPC-Bereich unterscheiden, werden verfeinert und weiterentwickelt.

- Im Rahmen eines Begleitprojekts zur Kompetenzentwicklung werden die für diese Analysysteme notwendigen Kompetenzen (z.B. spezielle Mathematik/Statistik-Kompetenzen als Ergänzung zu Datenmanagementkompetenzen aus der Informatik) erweitert.

V.4 Etablierung einer Datenföderation in Baden-Württemberg

Im Rahmen der Forschung werden zur Datenauswertung und -verarbeitung die wissenschaftlichen Daten regelmäßig zwischen unterschiedlichen (Speicher-)Systemen transferiert. Für den effizienten Datentransfer sind nicht nur die großen Datenmengen selbst, sondern auch die oft nur für geringe Latenzen und kurze Wege ausgelegten Protokolle für höchste Datenraten eine Herausforderung. Es müssen daher im Rahmen der Datenföderation nicht nur verschiedene Zugriffsmechanismen und besondere Anbindungen auf Netzebene, sondern auch intelligente Verteilungsmechanismen umgesetzt werden, um ggf. mit lokalen oder regionalen Datacaches eine möglichst effiziente Anbindung an die Analyse-/Rechensysteme zu realisieren.

bw-* Innovationsprojekte

Ein Kernelement zur Umsetzung universitätsübergreifender Dienste in Baden-Württemberg sind die **bw-* Innovationsprojekte**²³ und die auf die Bedürfnisse der Universitäten im Land abgestimmte Netzinfrastruktur **BelWü**²⁴.

Neben der breitbandigen Konnektivität von derzeit 100 Gbps unter den Universitäten in Baden-Württemberg und mit dem Internet-Backbone werden durch BelWü im Rahmen von einigen dieser Innovationsprojekte und betriebsnahen Vorhaben **spezielle Lösungen** umgesetzt, die insbesondere die für die Leistungsfähigkeit kritische **geringe Latenz** realisieren.

Andere Basisinfrastrukturdienste wie **bwIDM** zur Föderierung der lokalen Identitäten an den Einrichtungen oder Begleitprojekte wie **bwHPC-S5** realisieren unter Einbindung der lokalen Infrastruktureinrichtungen kontinuierliche Weiterentwicklung der Dienste und Prozesse.

Hinzu kommen neue Projekte zur Etablierung zukünftiger landesweiter Dienste. Als Beispiel bestehender Projekte können **bwVisu**, **bwCloud** und **bwDataArchiv** genannt werden.

Die Basis für diese BaWü-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

²³ <https://www.alwr-bw.de/kooperationen>

²⁴ <http://belwue.de>

Systemen, die durch die jeweiligen Fachcommunities 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 Communities 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.

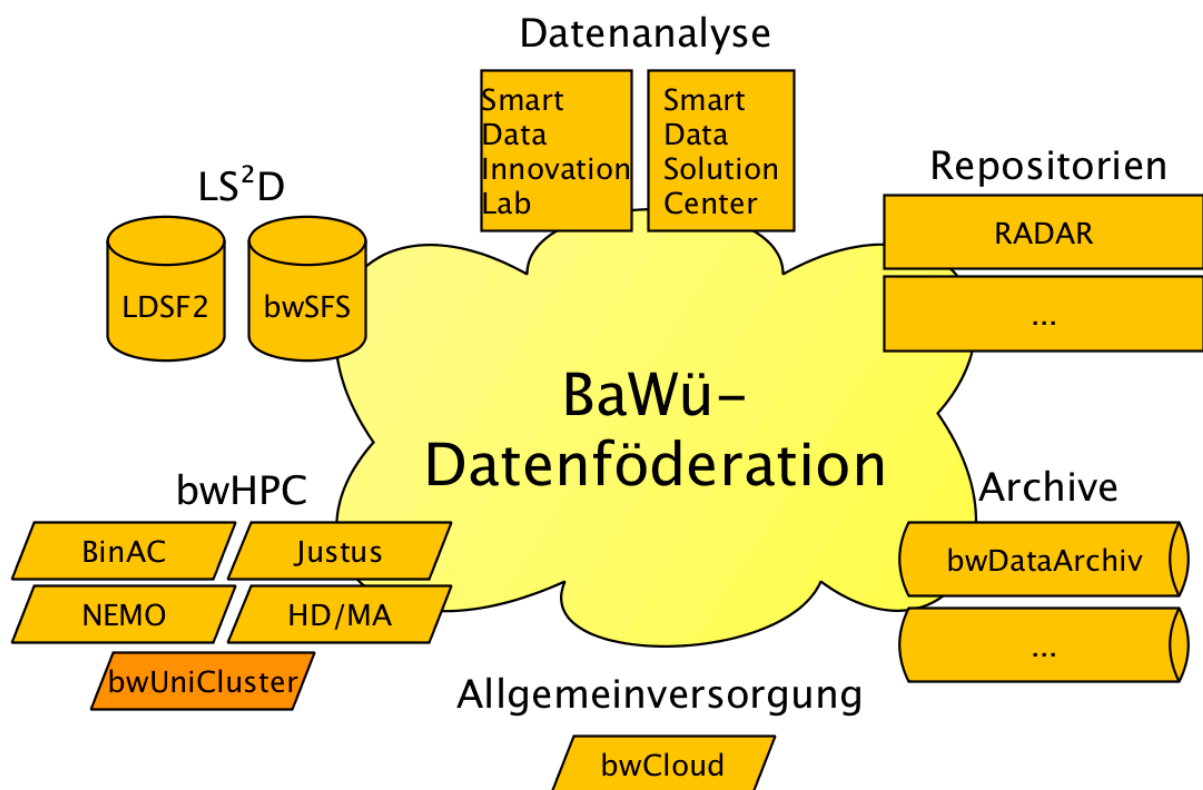


ABBILDUNG 5: AUFBAU BAWÜ-DATENFÖDERATION

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 und künftig evtl. EOSC) und internationaler Ebene (Research Data Alliance), Open Source, aber auch auf Ergebnisse der aktuell vom

Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg geförderten E-Science-Projekte.²⁵

Folgende Anpassungen an die landesspezifischen Aspekte und Anforderungen sind noch im Rahmen eines entsprechenden Entwicklungsprojektes durchzuführen:

- 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 BaWü-Datenföderation dar. Dazu sind analog zu den Kompetenzzentren auf dem HPC Tier-3 zur Nutzerunterstützung im Bereich Datenmanagement Data-Life-Cycle-Labs zu realisieren und wie nachfolgend beschrieben in die Gesamt-Supportstrukturen zu integrieren.
- Die Betriebsmodelle werden weiterentwickelt, damit sie einen für Nutzer transparenten Wechsel zwischen einzelnen Systemen gewährleisten und eine nahtlose Integration von nutzerfinanzierter projektspezifischer Hardware im Regelfall ermöglichen können.

Die Etablierung der BaWü-Datenföderation soll aufgrund der Vielfalt der Aufgaben im Begleitprojekt bwHPC-S5 als neuer Aufgabenbereich aufgenommen werden. Die Umsetzung der BaWü-Datenföderation soll in mehreren Stufen erfolgen:

- **Stufe 1 (Ausbau und Etablierung bestehender Methoden):** Unter Nutzung der bestehenden Möglichkeiten von bwIDM und unter Nutzung IP-basierter Standardprotokolle werden in einem ersten Schritt die LSDF-Infrastruktur, bwDataArchiv und die Speichersysteme der bwForCluster und des bwUniCluster angebunden. Dabei wird auch die Anbindung von Speichersystemen von Tier-2- und Tier-0/1-Systemen untersucht. Es steht dabei keine hochperformante Anbindung im Vordergrund, sondern primär die Vernetzung und Anbindung. Ziel dieser Stufe ist, neben den aktuellen und künftigen Anforderungen an bwIDM auch die Limitierungen auf Netz- und Protokollebene zu analysieren.
- **Stufe 2 (Verbesserung und Ertüchtigung der technischen Infrastruktur):** Zeitlich überlappend mit den oben beschriebenen Aktivitäten werden für ausgewählte Systeme Mechanismen zur intelligenten und automatisierten Replikation (Caching), Etablierung hierarchischer Speicherstrukturen, die z. B. selten genutzte Daten automatisiert auf geeignete Speichersysteme migrieren, und Lösungen für schnelle Austauschmechanismen untersucht und in einzelnen Bereichen implementiert. Dazu gehören auch dedizierte Netzanbindungen über Einrichtungsgrenzen hinweg. Diese Lösungen werden dem Bedarf entsprechend ausgebaut, im Gegensatz zu Stufe 1 wird eine Vollvermaschung zunächst nicht angestrebt. Neue Technologien in der föderierten Umgebung wie Objektspeicher werden erprobt und gegebenenfalls eingeführt.

²⁵ <http://www.forschungsdaten.info>

- **Stufe 3 (Verbesserung und Ertüchtigung der Software-Komponenten):** Während in den ersten beiden Ausbaustufen eine Fokussierung auf die Infrastruktur erfolgt, werden in Stufe 3 darüber hinausgehende Software-Komponenten etabliert, die Daten nicht nur in ihrer rohen Form übertragen lassen, sondern diese im Sinne eines (durchgehenden) Forschungsdatenmanagements vor einer Übertragung geeignet konvertieren, reduzieren oder verändern. Dies umfasst neben offensichtlichen Methoden wie Anonymisierung, Pseudonymisierung und Datenaggregation auch die Möglichkeit, Datenanalysen am Speicherort der Primärdaten durchzuführen und nur Ergebnisdaten mit beschreibenden Metadaten zu transferieren. Diese Ausbaustufe erfordert gemeinsame Innovationsprojekte von Infrastruktur- und Anwendungspartnern, die ausgehend von konkreten Anwendungsfällen geeignete Mechanismen entwickeln und erproben. Für diesen Bereich ist mittel- und langfristig auch zusätzliches Personal über die Projektmitarbeiter hinaus erforderlich, dass die notwendige Beratung der Wissenschaftler sicherstellt.
- **Stufe 4 (Konsolidierung und Erweiterung der Stufen 1-3):** Langfristig wird die in den Stufen 1 bis 3 gewonnenen Erkenntnisse und Ergebnisse evaluiert. Darauf aufbauend wird die Basisinfrastruktur für die Grundanforderungen des Datenmanagements um zusätzliche Funktionalitäten wie virtuelle Forschungsumgebungen oder umfassende Workflowsteuerung erweitert. Angestrebt wird eine möglichst nahtlose Integration in nationale und internationale Datenverbünde wie eine künftige NFDI.

V.5 Weiterentwicklung der Support-Strukturen

Weiterentwicklung des erweiterten Begleitprojektes zur Landesstrategie (bwHPC-S5)

Kernelement des neu ausgerichteten, erweiterten Begleitprojektes bwHPC-S5 Scientific Simulation und Storage Support Service zur Landesstrategie 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 aufgrund der Erfahrungen die Ausweitung des landesweiten HPC-Schulungsprogramms unter Berücksichtigung von DIC und LS²DM. Bei allen Projektzielen wird verstärkt auf die Abstimmung mit den Tier-1- und Tier-2-Kompetenzen geachtet.

Das Projekt ist in drei Bereiche gegliedert. Der erste Bereich, Nutzerunterstützung, übernimmt die fachlich ausgerichtete und auf Kompetenzzentren basierende Support-Struktur aus bwHPC-C5 und erweitert diese um die Themen DIC und LS²DM. Das Ziel ist eine integrierte Unterstützung über alle Themengebiete hinweg. Für den Bereich des abgestimmten landesweiten Betriebs der Infrastruktur wird die bisherige Abstimmung in den jeweiligen Bereichen HPC und Daten noch enger verzahnt und aufeinander abgestimmt. Im Bereich *Neue Technologien* werden mögliche zukünftige Dienste gemeinsam prototypisch umgesetzt und evaluiert. Die sich ändernden Anforderungen der Nutzer, die Erfahrungen in der Nutzerunterstützung und die Ergebnisse der Technologieevaluation bilden die Grundlage für die Weiterentwicklung des Angebots in allen Bereichen.

Ausbau der Kompetenzzentren

Im neu ausgerichteten und erweiterten Begleitprojekt soll die föderative Wissenschaftsunterstützung zunächst um folgende drei fachliche Schwerpunkte ausgebaut werden:

- Ingenieurwissenschaften durch das HLRS in Stuttgart (Tier-0/1), Karlsruhe (Tier-2) und Esslingen (Tier-3)
- Geowissenschaften, federführend durch Tübingen und Hohenheim
- Global Systems Science am HLRS in Stuttgart

Für jeden der Schwerpunkte wird im Bereich HPC ein Kompetenzzentrum etabliert, wobei für die Ingenieurwissenschaften die notwendigen Strukturen am HLRS und SCC schon existieren und nur eine Neueinrichtung in Esslingen notwendig ist. Ebenso existieren bereits die Strukturen für die Geowissenschaften in Tübingen und Hohenheim.

Integration von Nutzerunterstützung zu den Datensystemen in die Kompetenzzentren

Die im Rahmen des bisherigen bwHPC-C5 Projekts entstandenen Supportstrukturen müssen im anschließenden Begleitprojekt bwHPC-S5 durch die Erfahrungen und Erkenntnisse aus den Data Life Cycle Labs²⁶ erweitert werden, um die landesweite Benutzerunterstützung einer BaWü-Datenföderation sicherzustellen und um die Integration von HPC, DIC und Datenmanagement wirkungsvoll vorantreiben zu können. Die Data Life Cycle Labs bieten Unterstützung bei der Optimierung des kompletten Lebenszyklus großvolumiger wissenschaftlicher Daten von der Erzeugung in Experimenten und Simulationen bis hin zur Archivierung. Diese Expertise wird einen wichtigen Beitrag bei der fachspezifischen Unterstützung der Kompetenzzentren leisten.

VI. Governance

VI.1 Bestehende Governance

Die wissenschaftsgeleitete Steuerung der Nutzung und Weiterentwicklung dieser Infrastruktur wird gewährleistet durch eine übergreifende und in den letzten Jahren etablierte Governance-Struktur.²⁷ 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)

²⁶ <https://www.scc.kit.edu/ueberuns/8057.php>

²⁷ vgl. hierzu Wiebelt Bernd, Janczyk Michael, von Suchodoletz Dirk, Aertsen Ad, Rotter Stefan, Quast Günter, Schumacher Markus, Greiner Andreas, „Strukturvorschlag für eine bwHPC-Governance der ENM-Community“ bzw. Wesner Stefan, Walter Thomas, Wiebelt Bernd, von Suchodoletz Dirk, Schneider Gerhard, „Strukturen und Gremien einer bwHPC-Governance Momentaufnahmen und Perspektiven“ In: Kooperation von Rechenzentren Governance- und Steuerung - Organisation, Rechtsgrundlagen, Politik 2016, de Gruyter

- bwDATA-Steuerungskreis

Bewusst sind die einzelnen Kreise so konstituiert (s. u.), dass eine personelle Überschneidung und damit ein direkter Informationsfluss und eine Durchlässigkeit der einzelnen Ebenen gewährleistet sind.

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. Die technische Abstimmung zwischen den beiden Zentren findet im Rahmen der Kooperation in SICOS BW GmbH²⁸ statt.

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.

Dem bwHPC-Lenkungskreis sind folgende Themen zur Beschlussfassung vorzulegen:

- bwHPC Konzept und seine Fortschreibungen,
- bwHPC Evaluierungen,
- Konzept für die Erhebung des Bedarfs und dessen Ergebnisse,
- Konzepte und Anträge für Beschaffungen im Rahmen von bwHPC,

²⁸ SICOS BW GmbH (SIMulation, COmputing, Storage), <https://www.sicos-bw.de/>, gemeinsam von Universität Stuttgart und KIT gegründet und vom Land gefördert

- Betriebskonzepte für bwHPC,
- Projektanträge, die Begleitprojekte für bwHPC anstreben.

Der **bwDATA-Steuerungskreis**²⁹ steuert 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
 - o der Pädagogischen Hochschulen
 - o des Landesarchivs
 - o der Landesbibliotheken
 - o der Kunst- und Musikhochschulen
 - o der vier Universitätsklinika

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 Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg als ständiger Gast teil.

VI.2 Fortführung der Governance

Die oben beschriebene und in den Konzepten bwHPC und bwDATA formulierte Governance-Struktur hat sich aus Sicht der Beteiligten bewährt und ihre Aufgaben erfüllt. Sie soll nun 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.

VII. Transfer in die Wirtschaft

Neben der Etablierung der Infrastruktur und der effizienten und qualitativ hochwertigen Unterstützung der Wissenschaftler werden bereits heute die Möglichkeiten zum Wissenstransfer von

²⁹ bwDATA - Rahmenkonzept der Hochschulen des Landes Baden-Württemberg für datenintensive Dienste bwDATA (2015-2019): <http://dx.doi.org/10.15496/publikation-21187> und Umsetzungskonzept der Universitäten des Landes Baden-Württemberg für datenintensive Dienste – bwDATA Phase I (2013-2014): <http://dx.doi.org/10.15496/publikation-21188>

der Wissenschaft in Wirtschaft und Gesellschaft intensiv genutzt. Dabei ist es hier nicht nur das Ziel, den Wirtschaftsstandort Baden-Württemberg nachhaltig und langfristig zu fördern, sondern in gleichem Maße sollen auch die Anforderungen und Problemstellungen der Industrie zurück in die Wissenschaft gespiegelt werden.

VII.1 Ressourcennutzung durch die Industrie

Sowohl das Tier-1-Rechnersystem am HLRS wie auch die Datenanalysesysteme und Rechnerressourcen am SCC werden über die hww³⁰ und SICOS BW im Land für die Wirtschaft zugänglich gemacht. So wird die Bereitstellung von Rechenzeit für industrielle Nutzer seit 1996 über ein Public-Private-Partnership zwischen Porsche und T-Systems einerseits und dem KIT und der Universität Stuttgart sowie dem Land Baden-Württemberg andererseits realisiert. Die hww GmbH vertreibt derzeit jährlich etwa 100 Millionen Prozessorkernstunden an die heimische Industrie. Die Einnahmen aus dem Rechenzeitverkauf werden entsprechend den von den Zuwendungsgebern festgelegten Regularien an den beiden Standorten (HLRS und SCC) reinvestiert.

VII.2 Solution Center

Um den Know-How-Transfer in die Industrie zu ermöglichen und das Zusammenspiel von Forschung und Industrie im Bereich der Daten zusammen mit HPC zu verbessern, wurde das Konzept der Solution Centers entwickelt. Derzeit existieren drei Solution Centers:

1. Automotive Simulation Center Stuttgart (ASCS):³¹ Das Automotive Simulation Center Stuttgart wurde 2008 aufgrund einer Initiative des Landes Baden-Württemberg, der Firmen Porsche und Daimler sowie des HLRS als erstes Solution Center in der Landesstrategie gegründet. Es fokussiert inhaltlich auf die Themen:
 - Elektromobilität
 - Fahrzeugbau
 - Informationstechnische Methoden
 - Numerische Verfahren

Das Zentrum hat mittlerweile 33 Mitglieder aus dem Bereich Fahrzeughersteller, Rechnerhersteller, Softwarehersteller, Service Provider sowie aus der Forschung. Insbesondere die Mitgliedschaft von Porsche, Daimler, Ford und Honda zeigt die große Breitenwirkung des Zentrums sowie seine internationale Ausstrahlung.

³⁰ Höchstleistungsrechner für Wissenschaft und Wirtschaft GmbH, <http://www.hww.de>

³¹ <https://www.asc-s.de>

2. Energy Solution Center (ENSOC):³² Das im Jahr 2009 gegründete Energy Solution Center (EnSoc) führt anwendungsorientierte Forschung in Energiewirtschaft und Energietechnik unter Einsatz von HPC-Technologien durch. Es bündelt dazu die inhaltlichen Kompetenzen seiner Mitglieder zu Energiewirtschaft, Energietechnik, IT, Lösungsmethoden und der Durchführung von Forschungsprojekten. Mitglieder des EnSoC sind das KIT, die EnBW Energie Baden-Württemberg AG, die Siemens AG, die SAP AG, die T-Systems Sfr GmbH und die Hewlett-Packard Deutschland GmbH.
3. Smart Data Solution Center Baden-Württemberg (SDSC-BW):³³ Im Rahmen des SDSC-BW unterstützen KIT/SCC und die SICOS-BW GmbH gemeinsam kleine und mittelständische Unternehmen in Baden-Württemberg bei der Hebung von Wertschöpfungspotentialen mittels innovativer Smart-Data-Technologien. SDSC-BW stützt sich auf die Forschungsplattform „Smart Data Innovation Lab“ (SDIL), die in Zusammenarbeit zwischen Forschung und Industrie konzipiert wurde und am SCC betrieben wird. Neben dem KIT zählen Konzerne wie Bayer, Bosch, IBM, SAP, Siemens und die Software AG ebenso zu den Gründungspartnern wie das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI), die Fraunhofer-Gesellschaft und das Forschungszentrum Jülich.

Weitere Solution Center sind mit Unterstützung des HLRS in Vorbereitung:

1. Media Solution Center: Das Media Solution Center ist eine Initiative, die auf den großen Bedarf der Medienwelt an Simulation zurückgeht. Seit 2015 läuft ein Vorbereitungsprojekt, das 2018 durch eine Vereinsgründung abgeschlossen wird. Initiatoren des Media Solution Center sind das HLRS, die Hochschule für Medien, die Filmakademie Ludwigsburg sowie das Zentrum für Kunst und Medien in Karlsruhe. Dadurch wird das Media Solution Center sowohl inhaltlich als auch geographisch eine große Breite im Land abdecken können. Inhaltlich wird sich das Zentrum auf folgende Themen fokussieren:
 - Simulation in der Kunst
 - Simulation im Film (Animation, special effects)
 - Simulation erweiterter und virtueller Realitäten
2. Medical Solution Center: Das Medical Solution Center wird ab 2018 als Pilotprojekt vorangetrieben um das Potential der Simulation im Medizinbereich zu bündeln. Unter der Leitung des HLRS werden verschiedene Unternehmen der Medizinbranche aus Baden-Württemberg Themen und Herausforderungen ausarbeiten. Darüber hinaus muss eine Adaption des Konzepts der Solution Centers für den medizinischen Bereich erarbeitet werden. Insbesondere die Problematik der Vertraulichkeit und Patientensicherheit muss hier im Vordergrund stehen. Die Gründung des Medical Center ist für 2021 geplant.

³² <http://www.ensoc.de>

³³ <http://www.smart-data-solution-center.de>

VIII. Glossar

ALWR-BW	Arbeitskreis der Leiter und Leiterinnen der wissenschaftlichen Rechenzentren bzw. Informationszentren Baden-Württembergs
ASCS	Automotive Simulation Center Stuttgart
BaWü-Datenföderation	Baden-Württemberg Datenföderation
bwHPC-C5	Coordinated Compute Cluster Competence Centers
bwHPC-S5	Scientific Simulation und Storage Support Service
bwSFS	Storage for Science
de.NBI	German Network for Bioinformatics Infrastructure
DIC	Data Intensive Computing
DMP	Daten-Management-Plan
ENSOC	Energy Solution Center
FAIR	Findable, Accessible, Interoperable, and Reusable
FDM	Forschungsdatenmanagement
FDMI	Forschungsdatenmanagement-Infrastruktur
GA	Gauß-Allianz
HPDA	High Performance Data Analytics
HPC	High Performance Computing
IWR	Interdisciplinary Center for Scientific Computing
LNA-BW	Landesnutzerausschuss
LSDF	Large Scale Data Facility
LS ² DM	Large Scale Scientific Data Management
NFDI	Nationale Forschungsdaten-Infrastruktur
RADAR	Research Data Repository
RfII	Rat für Informations-Infrastrukturen
SCC	Steinbuch Centre for Computing
SDSC-BW	Smart Data Solution Center Baden-Württemberg
SDIL	Smart Data Innovation Lab
SDS@hd	Scientific Data Storage
SICOS-BW	Simulation Computing Storage, GmbH