

Vorwort

Die nun fünfte Ausgabe des bwHPC-Symposiums zeigt, dass sich die Veranstaltung fest etabliert hat. Sie ist Plattform für den Austausch von Wissenschaftlerinnen und Wissenschaftlern mit den Betreibern der großen, förderierten Forschungsinfrastrukturen. Das Symposium versucht, eine Brücke zwischen den verschiedenen Fachdisziplinen und den technisch-organisatorischen Aspekten zu schlagen. Es ergänzt die etablierte Governance aus Landesnutzerausschuss und den lokalen Gremien. Dem diesjährigen Symposium fiel die zusätzliche Rolle zu, die erste Projektphase von bwHPC-S5 einzuleiten, nachdem das fünfjährige Vorgängerprojekt bwHPC-C5 abgeschlossen wurde.¹

Erfolgreiche Wissenschaft benötigt leistungsfähige Infrastrukturen: Forschung bildet eine zentrale Säule im Selbstverständnis moderner Gesellschaften. Digitalisierte Arbeitsprozesse prägen alle Wissenschaftsdisziplinen. Die Stärke eines Wissenschaftsstandorts leitet sich wesentlich von der Verfügbarkeit attraktiver, integrierter und skalierender Forschungsinfrastrukturen ab. Die Computational Science und damit HPC-Systeme als deren technisches Fundament gewinnen unablässig an Bedeutung. Deshalb besteht eines der Ziele der im Juli 2018 gestarteten bwHPC-S5 Begleitaktivitäten darin, die Forscher*innen in ihrer Arbeit mit den leistungsfähigen Infrastrukturen umfassend zu unterstützen, ohne sie jedoch aus der individuellen Verantwortung für ihre Daten und wissenschaftlichen Workflows zu entlassen.

Der Erkenntnis folgend, dass heutige Anforderungen nicht mehr sinnvoll von einzelnen Universitäten oder Forschungsinstitutionen bedient werden können, koordinieren die wissenschaftlichen Rechenzentren des Landes Baden-Württemberg ihre Aktivitäten im Bereich HPC und der Speicherung großer Datenmengen unter

¹Vgl. hierzu <http://www.bwhpc.de/projektaufgaben.php>

dem gemeinsamen Dach des »Umsetzungskonzept(es) der Universitäten des Landes Baden-Württemberg für das High Performance Computing (HPC), Data Intensive Computing (DIC) und Large Scale Scientific Data Management (LS²DM)«. Damit baut das Land ein wesentliches Alleinstellungsmerkmal bei der Unterstützung der Wissenschaften aus und strebt eine Vorreiterrolle bei der Etablierung von Nationalen Forschungsdateninfrastrukturen (NFDI) und Science Data Centern an.

Der föderierte Betrieb großer Forschungsinfrastrukturen verändert das Umfeld und die Anforderungen der beteiligten Universitätsrechenzentren. Sie müssen sich verstärkt abstimmen und gemeinsame Lösungen finden. Nutzer der Infrastrukturen kommen nun nicht mehr allein aus der eigenen Einrichtung, sondern aus dem ganzen Land. Umgekehrt müssen Forscher*innen der eigenen Universität dabei unterstützt werden, Ressourcen an anderen Standorten zu verwenden. Leistungserbringung, Kosten und Zugehörigkeit zu einer Forschungseinrichtung fallen nicht mehr zwangsläufig zusammen, was neue Herausforderungen in der Leistungsverrechnung und Finanzierung mit sich bringt.

Mit der bereits vor mehr als zehn Jahren mit dem bwGRiD gestarteten HPC-Initiative des Landes wurde eine neue Kultur der Bereitstellung und Nutzung großer Forschungsinfrastrukturen eingeleitet. Durch die Spezialisierung der HPC-Systeme und die Bündelung gemeinsamer Anforderungen der verschiedenen Forschungsgruppen wurden Anstrengungen und Anschaffungen zusammengeführt. Diese nutzt nicht nur Economies-of-Scale, die Erkenntnis, dass der Aufwand für das Design, die Beschaffung und den Betrieb eines Clusters mit zunehmender Größe fast nicht mehr ansteigt. Es erlaubt ebenfalls eine deutlich effizientere Auslastung der Ressourcen, da üblicherweise in Einzelsystemen anfallende Leerlaufzeiten durch andere Nutzer gefüllt werden. Damit entstehen klare Mehrwerte, die es erlauben, einige deutlich größere Systeme mit hoher Leistung zu gleichen Kosten wie viele kleine mit deutlich geringerer Gesamtleistung zu beschaffen. Dabei erlauben es Mechanismen wie »Fairshare« Nutzungsanteile am Gesamtsystem in beliebiger Stückelung herauszugeben. Dieses senkt die Einstiegsschwelle insbesondere für kleinere Gruppen und Nachwuchsforscher*innen. Für große Gruppen und traditionelle Großantragsteller besteht der Mehrwert in einem zügigen Start ihrer Forschung, einer potenziell noch größeren Ressource als in einer individuellen Beschaffung ohne die Nachteile der Aufwendungen für eine eigene Betriebsmannschaft. Durch eine gestaffelte Beschaffung und regelmäßige Erneuerungszyklen bleibt die Attraktivität der Ressourcen

langfristig erhalten, und die Beteiligung neuer Nutzergruppen kann sichergestellt werden. Das verändert die Landschaft großer Forschungsinfrastrukturen nachhaltig.

Die im »High Performance Computing« und »High Throughput Computing« zunehmend genutzten großen Datenmengen in stärker föderierten Forschungsinfrastrukturen erfordern geeignete Strategien für ihre Haltung und ihren Zugriff, da selbst bei hochperformanten Netzwerken das Kopieren von Datensätzen immer weniger eine effiziente Lösung darstellt. Im Zuge der Entstehung, Verarbeitung und nachhaltiger Langzeitlagerung und Nachnutzung ändern sich die Anforderungen an das Datenmanagement bezüglich Zugriffsrechten, Leistungserwartungen und Absicherungen im Zeitablauf. Hierzu zählt die zunehmend komplexe Datenhaltung, die von der Generierung der Daten in Messinstrumenten, Simulationen und vielfältigen weiteren Verfahren über das Pre- und Postprocessing bis zur Verarbeitung und Visualisierung in Spezielsystemen und der langfristigen Haltung und Publikation reicht. Im Lebenszyklus von Forschungsdaten entstehen bei den Forschenden verschiedenartige Anforderungen, die von der schnellen Speicherung bei der Datenerhebung, über die Verarbeitung in HPC- und Cloudsystemen bis hin zur Visualisierung reichen. Hinzu kommt die notwendige Aufarbeitung der Daten im Sinne »guter wissenschaftlicher Praxis«, deren Langzeitspeicherung und -verfügbarkeit für Publikation und Nachnutzung. Das Forschungsdatenmanagement (FDM), die nachhaltige und zukunftsorientierte Speicherung von Forschungsdaten, ihre Verfügbarmachung und optimalerweise Publikation, rückt damit in den Fokus moderner Forschungsprozesse. Nachnutzung und Reproduzierbarkeit von Forschungsergebnissen benötigen sowohl Verständnis und Bereitschaft der Forscher*innen, als auch neuartige Dienstleistungen seitens der zentralen Infrastrukturanbieter.

Langfristig müssen geeignete Maßnahmen und Workflows umgesetzt werden, um den Wissenschaftler*innen nicht nur einen Großteil der Datenmanagementaufgaben abzunehmen, sondern auch die Effektivität der Ressourcennutzung zu steigern. Dies ist insbesondere bei der Nutzung von HPC-Ressourcen wünschenswert, indem der dort stattfindende Batch-Betrieb durch die automatische Versorgung mit Daten optimiert wird. Neben weiteren »lokalen« Softwarekomponenten, wie beispielsweise Versionsmanagement, ist auch der Anschluss an, beziehungsweise die einfache Nutzbarkeit bestehender externer Forschungsdateninfrastrukturen und der zukünftigen Nationalen Forschungsdateninfrastruktur notwendig, um Wissenschaft-

lern die Nutzung der fachspezifischen Repositorien zu ermöglichen. Die Aufgabe der Betreiber und des bwHPC-S5-Projekts besteht demnach zusätzlich darin, beständig neue Lösungen zu suchen und die föderativ angelegten Forschungsinfrastrukturen im Sinne der Forschenden beständig weiterzuentwickeln. Die Anwendung von Virtualisierungs- und Containerisierungsstrategien ist nur ein Beispiel hierfür.

Der Tagungsband besteht aus drei Teilen, wobei im ersten Teil die konzeptionellen und projektbezogenen Darstellungen im Vordergrund stehen. Der zweite Teil beinhaltet die durch ein Peer-Review begutachteten Beiträge der Forscher*innen, die eine Verbindung von Forschungsfragestellungen mit konkreten Berechnungen auf den jeweiligen bwHPC-Systemen herstellen. Im dritten Teil finden sich die durch die Projektpartner begutachteten Texte zu administrativen und betrieblichen Fragestellungen und Überlegungen zur technischen Weiterentwicklung der Betriebsmodelle.

Die in diesem Tagungsband zusammengefassten Konzepte und Aktivitäten beziehen zu großen Teilen ihre Kraft aus der langfristig angelegten Unterstützung des Ministeriums für Wissenschaft, Forschung und Kultur in Baden-Württemberg. Hierzu zählen zuvorderst die finanziellen Beiträge für die Beschaffung und regelmäßige Erneuerung der HPC-Systeme als auch die Förderung der bwHPC-C5- und bwHPC-S5-Begleitprojekte. Weiterhin profitieren die beteiligten Forschungseinrichtungen durch die eScience-Initiative des Landes, durch die beispielsweise das ViCE-Projekt seine Unterstützung erfuhr.

Die Herausgeber
Freiburg, April 2019