

Molekül- und Materialforschung: Daten leicht teilen

„Science Data Center für Molekulare Materialforschung“ entwickelt Digitalisierungsbausteine für wissenschaftliche Daten – von Erfassung über Prozessierung bis zur öffentlichen Archivierung.



Daten aus vielen Laboren nachhaltig und gemeinschaftlich nutzen, ist das Ziel des neuen Science Data Centers MoMaF. (Foto: Laila Tkotz/KIT)

Das Internet bietet uns rund um die Uhr direkten Zugriff auf das Wissen der Welt. Eigene Projekte profitieren vom Know-how vieler Köpfe und können wiederum mit Interessierten geteilt werden. Gerade Forschende, deren Handwerk der Umgang mit Daten ist, streben einen freien Informationsfluss an. Für die in Laboren erzeugten Rohdaten ergeben sich jedoch einige Hürden beim Datenaustausch. Das „Science Data Center für Molekulare Materialforschung“ am Karlsruher Institut für Technologie (KIT) will das nun in Zusammenarbeit mit der Hochschule Karlsruhe und dem FIZ Karlsruhe ändern und erhält dafür eine Förderung von 2,5 Millionen Euro vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (MWK).

„Wir schaffen in Baden-Württemberg eine E-Science-Infrastruktur, die unseren Forschenden beste Voraussetzungen für neuartige wissenschaftliche Ansätze bietet. Data Science kombiniert die

Monika Landgraf
Pressesprecherin,
Leiterin Gesamtkommunikation

Kaiserstraße 12
76131 Karlsruhe
Tel.: +49 721 608-21105
E-Mail: presse@kit.edu

Weiterer Pressekontakt:

Kosta Schinarakis
Redakteur/Pressereferent
Tel.: +49 721 608-21165
E-Mail: schinarakis@kit.edu

Holger Gust
Hochschule Karlsruhe
Tel.: 0721 925-1016
E-Mail: pr@hs-karlsruhe.de

Techniken der Mathematik und Informatik mit dem Wissen über verschiedene Anwendungsfelder und öffnet das Tor zu neuen Erkenntnissen“, sagt Ministerin Theresia Bauer. „Für den Standort Baden-Württemberg sind daher wissenschaftsgetriebene Datenzentren wie das am KIT von herausragender Bedeutung.“

„Mit dem neu geschaffenen Science Data Center sorgen wir dafür, dass Wissen auch die Grenzen von Institutionen überwindet und übergreifend wirken kann“, sagt Professor Holger Hanselka, Präsident des KIT. „Auf dem gemeinsamen Weg zu Lösungen für die globalen Herausforderungen kommen wir dank des Science Data Centers schneller voran.“

„Forschung lebt von der Kooperation und dem Austausch, allerdings ist das Teilen von Daten technisch nicht immer einfach“, so Professor Oliver Kraft, Vizepräsident des KIT für Forschung. „Die Einrichtung des Science Data Centers für Molekulare Materialforschung ist ein wichtiger Schritt auf dem Weg, Forschungsdaten gemeinschaftlich und nachhaltig zu nutzen.“

Forschung beschleunigen

Daten sind teuer. Unmengen Arbeits- und Rechnerstunden sowie teure Apparate und Materialien sind der Preis für wissenschaftliche Erkenntnisse, der in der Regel mit öffentlichen Geldern bezahlt wird. Der Wert der erhaltenen Daten ist häufig unvergänglich und auch ältere Daten lassen sich für aktuelle Untersuchungen nutzen. Insbesondere neue Analysemethoden können den Erkenntnisgewinn aus diesen Daten und damit ihren langfristigen Nutzen maximieren.

Sowohl die systematische Datensicherung als auch deren nachhaltige Bereitstellung bilden wesentliche Erfolgskriterien in der Wissenschaft. Beides einzuhalten ist nicht immer einfach. So fehlen häufig effiziente Werkzeuge, um Daten auszutauschen, für andere nachvollziehbar zu strukturieren und mit Metadaten zu versehen. Gelegentlich sind die rechtlichen Rahmenbedingungen nicht klar. Manchmal scheitert es schlicht an der Datenmenge, die sich in vielen Projektjahren angehäuft hat und lange Zeit verfügbar sein soll. Das „Science Data Center für Molekulare Materialforschung“, kurz MoMaF, will nun Hemmschwellen reduzieren und für Chemiker und Materialwissenschaftler passende Prozesse und Werkzeuge entwickeln, um praxisnahe Lösungen für aktuelle Probleme im Forschungsdatenmanagement zu bieten.

„Die gemeinsame Nutzung von Daten in den Materialwissenschaften beschleunigt nationale sowie internationale Forschung und damit auch Innovation in zentralen Forschungsbereichen wie Energie und

Gesundheit“, erklärt Professorin Britta Nestler, die am KIT und an der Hochschule Karlsruhe aktiv ist. Seit 2016 forscht sie auch am Materialwissenschaftlichen Zentrum für Energiesysteme (MZE) des KIT. Bislang nutzen Arbeitsgruppen in den Fachbereichen von Molekülchemie bis hin zu den Materialwissenschaften meist individuelle Lösungen des Datenmanagements, was dazu führt, dass Verfügbarkeit und Sichtbarkeit von Forschungsergebnissen geringer sind. Die Ergebnisse anderer Forschungszweige können kaum genutzt werden, um schnellere und umfangreichere Studien durchzuführen. „Es fehlt bislang ein übergreifender, einheitlicher Werkzeugkasten für das Speichern, Verarbeiten und Kuratieren von Forschungsdaten, die KI-gestützte Analyse sowie die interdisziplinäre Nachnutzung. Und eine Institution, welche übergreifend Kompetenzen bündelt, und für alle nutzbar macht.“

Effizientes Forschungsdatenmanagement

„Mit dem MoMaF werden wir Bausteine für die Digitalisierung entwickeln, welche alle Phasen von der Generierung der Daten bis zu deren nachhaltiger Archivierung abdecken“, erläutert Professor Stefan Bräse vom Institut für Organische Chemie des KIT, der ebenfalls am MZE aktiv ist. Dieser Beitrag zur Digitalisierung stellt sicher, dass Daten über Moleküle – sowie deren Wechselwirkungen zur Beschreibung von Materialien – so abgelegt werden, dass diese übergreifend auffindbar, zugänglich, bedienbar und mehrfach nutzbar sind, also die sogenannten FAIR Data-Prinzipien (steht für *Findable, Accessible, Interoperable, Reusable*) beachtet werden. Dabei werden nicht nur disziplinspezifisch und interdisziplinär relevante Forschungsergebnisse gesichert, sondern insbesondere auch die zum Verständnis der Daten erforderliche Aufbereitung und Analyse unterstützt. Ziel ist die Bereitstellung einer Software-Infrastruktur, die es erlaubt, die grundlegenden und spezifischen Anforderungen an wissenschaftliche Datensicherung und effiziente Nachnutzung abzubilden. „Das MoMaF wird Schlüsselemente für das Forschungsdatenmanagement liefern, die bisher weder national noch international verfügbar sind.“

Das MoMaF kann auf ein am KIT bereits etabliertes Konzept zurückgreifen: auf ein elektronisches Laborjournal (ELN, Electronic Lab Notebook) mit dem angebundenes Forschungsdatenrepositorium (öffentlich zugängliches Datenarchiv) Chemotion für das Fach Organische Chemie. Das ELN bietet Funktionen zur fachspezifischen Erfassung, Organisation, Prozessierung und Verknüpfung von Forschungsdaten. Diese Funktionen bilden die Basis einer strukturierten Speicherung und Verwendung der erhaltenen Daten wie etwa einer Nachnutzung durch andere Forschende. Ein direkter Transfer der erhaltenen wissenschaftlichen Daten in das Forschungsdatenreposito-

rium Chemotion erlaubt weiterführend etwa die unterstützte Generierung von Metadaten und die automatisierte Registrierung von eindeutigen, permanenten Benennungen (Persistent Identifier, PID) zur Anbindung externer, fachspezifischer Datenbanken. Chemotion ist eine weltweit beispielhafte Software und wurde mit dem SPARC-Europe Open Data Champion Award 2017 ausgezeichnet. Der bisher entwickelte Quellcode von ELN und Repository wird nun mit passenden Modulen erweitert, um auch in den angrenzenden Fachbereichen Molekülchemie, makromolekulare Chemie und Oberflächenchemie sowie virtuelle Materialentwicklung eingesetzt werden zu können. Auch wird ein Empfehlungsdienst als Softwaresystem implementiert, welches durch den Einsatz von Methoden des Maschinellen Lernens auf Organisations- und Analyseebene unterstützt sowie Empfehlungen zu Erhebung, Verwahrung, Kuratierung und Nachnutzung gibt.

Forschung auf höchstem Niveau

Die Beteiligung der Rechenzentren des KIT und der Hochschule Karlsruhe sowie der Bibliothek des KIT an MoMaF sichert die Integration des Science Data Centers in die Forschungs- und Lehrstrukturen der Partner und bildet eine Brücke zu den bereits etablierten Serviceleistungen des KIT im Forschungsdatenmanagement. Das Steinbuch Centre for Computing des KIT betreibt das Rechenzentrum und bringt langjährige Erfahrung im Umgang mit großen wissenschaftlichen Datenmengen durch Projekte des KIT wie etwa GridKa, LSDMA, LSDF und bwDataArchiv ein. Diese Expertise wird durch das Institut für Angewandte Informatik und Formale Beschreibungsverfahren des KIT unterstützt.

MoMaF möchte einen Beitrag in nationalen und internationalen Initiativen leisten, etwa zur Einrichtung und Unterstützung von Forschungsdateninfrastrukturen wie der deutschen Nationalen Forschungsdaten Infrastruktur (NFDI) und der Europäischen Research Data Alliance. Die beiden Exzellenzcluster des KIT "3D Matter Made to Order" und "Energiespeicherung jenseits von Lithium" gehören zu den ersten Nutzern der Werkzeuge von MoMaF, ebenso soll es den Bedarf von weiteren Forschungsverbänden abdecken, etwa des SFB1176 und des SFB/TRR88, an denen Forschende des MoMaF ebenfalls beteiligt sind. Langfristig soll MoMaF als Forschungsinstrument eine zusätzliche Verankerung in den Programmen der Helmholtz-Gemeinschaft finden.

Das „Science Data Center für Molekulare Materialforschung“ wird zentral am und vom KIT betrieben und die entwickelte Software wird als Open Source einer breiten wissenschaftlichen Community zugänglich gemacht. Partner sind neben dem koordinierenden KIT die

Hochschule Karlsruhe – Technik und Wirtschaft sowie das FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur. Zur Evaluation der Nutzung an verschiedenen Hochschulstandorten soll die Infrastruktur zum Betrieb der ELNs auch an der Hochschule Karlsruhe aufgebaut werden. Das FIZ Karlsruhe analysiert rechtliche Handlungsfelder bei der Entwicklung und Implementierung der digitalen Bausteine.

MoMaF ermöglicht Forschenden in Baden-Württemberg, Forschung auf höchstem Niveau zu betreiben und sichert so die Wettbewerbsfähigkeit. Die Aufbauphase beträgt 4 Jahre. Das Budget von rund 3,5 Millionen Euro setzt sich aus rund einer Million Euro Eigenanteil und einer Förderung von 2,5 Millionen Euro durch das Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg zusammen.

Weiterführende Informationen:

Forschungsdatenmanagement des KIT: <https://www.rdm.kit.edu>

MZE: https://www.kit.edu/kit/pi_2016_161_materialwissenschaftliches-zentrum-am-kit-eroeffnet.php

Research Data Alliance: https://www.sek.kit.edu/kit_express_4016.php

ELN Chemotion: <https://chemotion.net/>

<https://openscholarchampions.eu/opendata/champion/datajournalsandrepositoriesgotogether/>

NFDI4Ing und NFDI4Chem: <https://www.tib.eu/de/service/aktuelles/detail/nationale-forschungsdateninfrastruktur-fuer-die-chemie-nfdi4chem/>



Als „Die Forschungsuniversität in der Helmholtz-Gemeinschaft“ schafft und vermittelt das KIT Wissen für Gesellschaft und Umwelt. Ziel ist es, zu den globalen Herausforderungen maßgebliche Beiträge in den Feldern Energie, Mobilität und Information zu leisten. Dazu arbeiten rund 9 300 Mitarbeiterinnen und Mitarbeiter auf einer breiten disziplinären Basis in Natur-,

Ingenieur-, Wirtschafts- sowie Geistes- und Sozialwissenschaften zusammen. Seine 25 100 Studierenden bereitet das KIT durch ein forschungsorientiertes universitäres Studium auf verantwortungsvolle Aufgaben in Gesellschaft, Wirtschaft und Wissenschaft vor. Die Innovationstätigkeit am KIT schlägt die Brücke zwischen Erkenntnis und Anwendung zum gesellschaftlichen Nutzen, wirtschaftlichen Wohlstand und Erhalt unserer natürlichen Lebensgrundlagen.

Diese Presseinformation ist im Internet abrufbar unter:
www.sek.kit.edu/presse.php

Das Foto steht in der höchsten uns vorliegenden Qualität auf www.kit.edu zum Download bereit und kann angefordert werden unter: presse@kit.edu oder +49 721 608-21105. Die Verwendung des Bildes ist ausschließlich in dem oben genannten Zusammenhang gestattet.

Mit seinem **Jubiläumslogo** erinnert das KIT in diesem Jahr an seine Meilensteine und die lange Tradition in Forschung, Lehre und Innovation. Am 1. Oktober 2009 ist das KIT aus der Fusion seiner zwei Vorgängereinrichtungen hervorgegangen: 1825 wurde die Polytechnische Schule, die spätere Universität Karlsruhe (TH), gegründet, 1956 die Kernreaktor Bau- und Betriebsgesellschaft mbH, die spätere Forschungszentrum Karlsruhe GmbH.