



Informatik als Treiber für E-Science

*Positionspapier der
Gesellschaft für Informatik e.V.*

Informatik - Grundlage für alle Wissenschaften

Albert Einstein nutzte eine Kreidetafel. Heutige Wissenschaftlerinnen und Wissenschaftler brauchen Informatik.

In den Erd- und Umweltwissenschaften liefern Satelliten und Sensornetzwerke kontinuierlich massive Datenströme. Von den Sentinel-Satelliten des Europäischen Copernicus-Programms werden beim Deutschen Fernerkundungsdatenzentrum (DFD) bis zu zehn Terabyte pro Tag archiviert und verarbeitet. Am Deutschen Klimarechenzentrum (DKRZ) wird durch Klimasimulationen ein jährliches Datenwachstum von etwa 75 Petabyte erzeugt.

Lebenswissenschaften wie Molekularbiologie, Biomedizin oder Lebensmitteltechnologie können durch die Untersuchung molekularer Prozesse auf der Basis von Genom-Daten ein neues Verständnis von Lebensvorgängen erreichen. Dies kann zu maßgeschneiderten Therapien in der Medizin führen oder die Züchtung robuster und ertragreicher Nutzpflanzen ermöglichen. Die dafür notwendige Datenmenge ist allein im European Bioinformatics Institute (EBI) von 15 auf 25 Petabyte im Jahr 2015 angestiegen.

Um die Datenmengen aktueller Forschung verarbeiten zu können, stellt z.B. die European Grid Infrastructure (EGI) 530.000 logische CPUs, 200 Petabyte Festplattenspeicher und 300 Petabyte Bandlaufwerke zur Verfügung. Diese Ressourcen werden gemeinschaftlich von 350 Rechenzentren in 56 Ländern der Welt betrieben. EGI ist nur möglich durch leistungsfähige Datenbanken, schnelle Rechnernetze und hohe Parallelität.

Die Ergebnisse der Informatik verändern in allen Wissenschaftsbereichen technologische Rahmenbedingungen und die Art und Weise wie Forschung durchgeführt wird. Sie verändern damit auch Kostenstrukturen und schaffen verbesserte Grundlagen für Forschung und Innovation. Studien zeigen, dass der Mehrwert einer effizienten IT-Infrastruktur um das Zwanzigfache höher sein kann als deren operationelle Kosten.¹

¹ <http://www.beagrie.com/static/resource/EBI-impact-report.pdf>



Warum ist E-Science wichtig?

E-Science beschreibt eine Vision des zukünftigen wissenschaftlichen Arbeitens und die mittel- bis langfristige Entwicklung der wissenschaftlichen Praxis dorthin. E-Science betrachtet im engeren Sinn die Anwendung von Informations- und Kommunikationstechnologie (IKT) in der modernen Wissenschaft und umfasst ein Spektrum von der Vorbereitung und Durchführung von Untersuchungen, über die Datenerfassung, Verarbeitung und Verbreitung von Ergebnissen bis hin zur Langzeitspeicherung und dem Zugang zu relevanten Materialien und Artefakten². Damit fördert E-Science Innovation in der kollaborativen, rechen- und datenintensiven Forschung in allen Disziplinen im gesamten Forschungslebenszyklus³. Dieser Transformationsprozess, die Umsetzung der Digitalen Agenda in der Wissenschaft, betrifft sowohl Natur- und Ingenieurwissenschaften als auch Sozial- und Geisteswissenschaften.

Ein wichtiger Treiber der Entwicklung zu E-Science sind die großen gesellschaftlichen Herausforderungen, wie Klimawandel und Energieversorgung, Gesundheit und demographischer Wandel oder der nachhaltige Umgang mit natürlichen Ressourcen. Von der Wissenschaft werden zunehmend konkrete Beiträge zum Umgang mit diesen Herausforderungen erwartet. Die Grundlagen dazu werden in global verteilten, kollaborativen und interdisziplinären Kooperationen erforscht, deren Erfolg von leistungsfähigen digitalen Arbeitsplattformen und einer dadurch möglichen integrierten Nutzung vorhandener Hard- und Software abhängig ist. Sensornetzwerke und Monitoringsysteme werden mit Hochleistungsrechnern und Speichersystemen durch leistungsfähige Middleware zu integrierten Forschungsplattformen verbunden. Die schnelle Analyse und Auswertung sehr großer experimenteller und empirischer Datenmengen schafft erst die Voraussetzungen, komplexe Zusammenhänge und Wechselwirkungen auf Systemebene zu erkennen und mit theoretischen Modellen abzugleichen. Diese Datenmengen müssen in global verteilten Wissenschaftsteams gemeinschaftlich interpretiert werden, um daraus neue Ergebnisse und Handlungsoptionen abzuleiten.

Die Fähigkeit zum Umgang mit der digitalen Transformation in Forschung und Lehre stellt eine Schlüsselkompetenz für unsere Universitäten, Forschungseinrichtungen und unsere Wissensgesellschaft insgesamt dar. E-Science setzt eine hochwertige technische Infrastruktur voraus. Die Förderung, Gestaltung und Umsetzung des Transformationsprozesses geht jedoch weit über die Bewältigung von technischen Herausforderungen hinaus. Das Themenfeld E-Science ist aktuell durch eine Vielzahl von nationalen und internationalen Aktivitäten gekennzeichnet. Die konkrete Umsetzung ist von spezifischen forschungspolitischen Rahmenbedingungen und komplementären Fördermaßnahmen abhängig⁴. Trotz der treibenden Kraft der Informations- und Kommunikationstechnologie für E-

² Bohle, S. (2013) "What is E-science and How Should it Be Managed?" Nature.com, Spektrum der Wissenschaft (Scientific American), http://www.scilog.com/scientific_and_medical_libraries/what-is-e-science-and-how-should-it-be-managed

³ IEEE International Conference on eScience, <https://escience-conference.org/>

⁴ Beispiele:

- DE: Rat für Informationsinfrastrukturen <http://www.rfii.de/de/index/>

- GB: JISC <https://www.jisc.ac.uk/>; Software Sustainability Institute <http://www.software.ac.uk/>

- NL: Netherlands eScience Center <https://www.esciencecenter.nl/>



Science wird die wesentliche Bedeutung der Wissenschaftsdisziplin Informatik für die Gestaltung und Umsetzung dieser, für den Forschungsstandort Deutschland wichtigen Entwicklungen noch nicht wahrgenommen.

Handlungsbedarf – Was ist zu tun?

Die Entwicklung von E-Science erfolgt aktuell in einer Vielzahl von Projekten, die von unterschiedlichen Organisationen im Rahmen von Programmen oder Einzelprojekten gefördert werden. Das Ziel, wissenschaftliche Daten und Software konsistent und nachhaltig zu verwalten sowie zwischen unterschiedlichen Fachdisziplinen austauschen und unabhängig von der Quelle nutzen zu können, wird durch fachbezogene Einzelentwicklungen nicht immer befördert. Gleiches gilt für Softwarewerkzeuge. Außerhalb des jeweiligen Faches sind diese Ressourcen nur mit erheblichem Aufwand nutzbar oder gar völlig unzugänglich.

Wissenschaftsdaten sind heterogen, dynamisch und folgen ebenso wenig einer strengen Normierungslogik wie interdisziplinäre Forschungsprojekte. Der Nutzen von Wissenschaftsdaten ergibt sich aus deren *fitness for use*, dem Nutzen für zukünftige Forschung, deren Anforderungen, z.B. an die Struktur der Daten, wir heute noch gar nicht kennen (können). Zur Prozessierung und zur Analyse von Daten, aber auch für die Modellierung und Simulation werden von Wissenschaftlern verbreitet Algorithmen und Programme entwickelt. Resultierende Softwarewerkzeuge werden nach Projektende häufig weder weiter verwendet noch weiter entwickelt. Ihr Entwicklungsaufwand ist aber meist genauso hoch wie der des wissenschaftlichen Experiments selbst. Während wir bei Mikroskopen, Teilchenbeschleunigern, Forschungsschiffen und anderen wissenschaftlichen Großgeräten eine möglichst vielfältige Nutzung bereits bei der Anschaffung einbeziehen, steckt dies bei wissenschaftlicher Software noch in den Kinderschuhen.

Als wissenschaftliche Fachdisziplin ist die Informatik dazu prädestiniert, derartige Entwicklungen aufzuzeigen und zugleich in ihrer Eigenschaft als konstruktive Disziplin belastbare Alternativen anzubieten. Kernthemen der Informatik waren und sind seit Jahren die Schaffung von Interoperabilität und die Verknüpfung heterogener Infrastrukturen. Konzepte und Architekturen zur Integration von Soft- und Hardware wurden ebenso von Informatikern entwickelt wie weltweite Standards zur Datenhaltung, zur Vernetzung und zum Management globaler Informationssysteme. Aus Sicht von Informatikerinnen und Informatikern ist es möglich, sich diesen Problemen wissenschaftlich zu nähern und bestehende Konzepte der Informatik und angrenzender Bereiche anzuwenden und weiter zu entwickeln. Es geht also nicht nur um eine technische Unterstützungsleistung für bestimmte Fachgebiete, sondern um die Chance, durch intensive Zusammenarbeit und gelebte Interdisziplinarität Fachsilos und Corpora miteinander zu vernetzen. Dies würde die Forschung sowohl im jeweiligen Fachgebiet als auch in der Informatik voranbringen. Daraus leiten wir Anforderungen an Wissenschaftspolitik, Universitäten und Forschungseinrichtungen sowie die Informatik selbst ab.



Herausforderungen für die Wissenschaftspolitik

Für die zielgerichtete Entwicklung von E-Science sind positive forschungspolitische Rahmenbedingungen notwendig. Die deutsche Wissenschaftspolitik sollte sich klar zu E-Science und dem koordinierten Aufbau einer neuen Generation von Infrastrukturen und Anwendungen für Forschung und Lehre bekennen. Diese Aktivitäten sollten grundsätzlich offenen Standards folgen, z.B. Zugang zu Daten und Software (Open Science) oder die Vernetzung von Daten (Linked Open Data). Aufbauend hierauf können fach- und domänenspezifische Standards und Konventionen etabliert werden, die jedoch mit anderen Disziplinen interoperabel bleiben müssen. Gerade im Hinblick auf zukünftige, interdisziplinäre Forschungsanstrengungen sollte daher eine gemeinsame methodische und technische Basis für einen nachhaltigen Zusammenhalt und Erweiterungsmöglichkeiten sorgen.

Die bereits existierenden, vielfältigen Wissenschaftsprogramme, welche auf die Digitalisierung einzelner Wissenschaften zielen, könnten durch eine solche Initiative gestärkt werden. Ein mögliches nationales Programm sollte u.a. folgende übergreifende Themen adressieren:

- Wissenschaftliches Datenmanagement
- Etablierung von Referenzarchitekturen und Standards
- Semantische Interoperabilität und Integration von Daten
- Software Engineering für wissenschaftliche Anwendungen
- Integrierte Modellierung und Systemwissenschaften, Data Science und Simulation
- Aufbau von Forschungsumgebungen mit kollaborativer Mensch-Maschine-Kooperation
- Einbeziehung von E-Science-Praktiken in die Graduiertenausbildung

Herausforderungen für Universitäten und Forschungseinrichtungen

Wissenschaft wird heute in globalem Kontext und internationalem Wettbewerb durchgeführt. Qualität und Innovationsgeschwindigkeit sind gleichermaßen wichtige Anforderungen. Neben aller Kreativität im Finden und Beantworten von Forschungsfragen entscheidet dabei auch der gekonnte Umgang mit Methoden und Werkzeugen einer digitalisierten und datenorientierten Wissenschaftswelt sowie die effiziente Vernetzung mit anderen global verteilten Wissenschaftlerinnen und Wissenschaftlern über Erfolg oder Misserfolg. Auch die Einbindung in und die Nutzung von nationalen und internationalen Forschungsinfrastrukturen^{5 6} hat hier einen zunehmenden Einfluss.

Universitäten und Forschungseinrichtungen müssen dabei zwei Entwicklungen beachten:

Der **Betrieb** eigener IT-Infrastrukturen ist eine Grundvoraussetzung für die aktuelle und zukünftige Wissenschaftspraxis. Das ist eine Herausforderung sowohl in finanzieller und personeller aber auch in konzeptioneller und organisatorischer Hinsicht. Rechenzentren und IT-Servicezentren sind prinzipiell unverzichtbare Dienstleister für die Wissenschaft und sollten die Sprache der Wissenschaft verstehen können. Sollten sie fachlich oder technisch dazu nicht fähig sein, werden die Fachbereiche das Heft des Handels selbst in die Hand nehmen und parallele Strukturen aufbauen, mit negativen Konse-

⁵ European Commission: http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=home

⁶ BMBF: <https://www.bmbf.de/de/roadmap-fuer-forschungsinfrastrukturen-541.html>



quenzen hinsichtlich interdisziplinärer und interinstitutioneller Kommunikation sowie Kosten, Sicherheit und Nachhaltigkeit. Der Schlüssel zum Erfolg liegt deswegen in einer zunehmenden Professionalisierung und größeren Anwendernähe zentraler Serviceeinrichtungen an Universitäten und außer-universitären Forschungsorganisationen.

Die **Entwicklung** neuer Wissenschaftsfelder benötigt zunehmend Kenntnisse in Werkzeugen und Methoden der Informatik. In vielen Wissenschaftsbereichen sind Projekte nur noch mit erheblichem IT-Aufwand erfolgreich durchführbar. Herausragende Kenntnisse in der Anwendung von Informatik auf die eigene Fachwissenschaft können ein Berufungskriterium für junge Wissenschaftlerinnen und Wissenschaftler werden. Für die Ausbildung junger Wissenschaftlerinnen und Wissenschaftler stellt E-Science einen notwendigen Baustein der zukünftigen Graduiertenausbildung dar, ebenso wie wissenschaftliches Publizieren oder Methodenkenntnisse. Die Ausprägung zukünftiger Formate der Graduiertenunterstützung erfordert neben der Bereitstellung von Werkzeugen auch die personelle Einbindung von Informatikkompetenzen. Einschlägig qualifizierte und interessierte Informatikerinnen und Informatiker an Universitäten sollten sich in eine solche Graduiertenausbildung einbringen.

Herausforderungen für Informatikerinnen und Informatiker

Informatikerinnen und Informatiker, in den Wissenschaftsdomänen arbeitende "Bindestrich-" Informatiker (z.B. Medizin-, Bio-, Wirtschafts-, Geo-Informatiker) sowie Domänenexpertinnen und -experten sind aufgefordert Plattformen zu schaffen, in denen übergreifende Fragen gemeinsam diskutiert und Lösungswege identifiziert werden können. Anfängliche Mehrinvestitionen rentieren sich durch höhere Nachhaltigkeit und Wiederverwendung gefundener Lösungen. Interessierte Informatikerinnen und Informatiker sollten sich an der Entwicklung offener Standards für den Austausch und die Verarbeitung wissenschaftlicher Daten beteiligen, die Interoperabilität mit anderen Fachdomänen in den Vordergrund stellen. Wir können und sollten als Wissenschaftlerinnen und Wissenschaftler in Deutschland nicht erwarten oder darauf vertrauen, dass eine einzelne große Institution im Alleingang Standards und Interoperabilität durchsetzt. Eine breit angelegte Koalition von Fachdomänen mit entsprechend breit gefächerten Fach- und IT-Kompetenzen scheint allerdings essenziell, um eine offene, leistungsfähige und effiziente wissenschaftliche Dateninfrastruktur zu erreichen.

Die Rolle der Gesellschaft für Informatik e.V.

Die Gesellschaft für Informatik steht bereit, einer solchen Entwicklung Raum zur Diskussion und Entfaltung zu bieten. Wir veranstalten Konferenzen und Workshops, die digitale und persönliche Plattformen der Kommunikation sein können. Wir führen gemeinsame Projekte mit der Wirtschaft, den Wissenschaften und der Gesellschaft durch. Wir sind Ansprechpartner für die Politik und Medien. Wir arbeiten zusammen mit anderen Fachgesellschaften, den Anforderungsgebern und Anwendern unserer Technologien. Wenn E-Science mehr sein soll als ein Begriff mit vielen Bedeutungen, sind Informatikerinnen und Informatiker der Schlüssel für fach- und domänenübergreifende Ansätzen zum Austausch und zur Analyse von Daten aus unterschiedlichen Bereichen anhand gemeinsam entwickelter Prozess- und Datenstandards. Die Gesellschaft für Informatik kann als Katalysator für entsprechende Projektansätze ihren Teil beitragen.