

Dr. Heike Neuroth
Leitung Abteilung Forschung & Entwicklung der SUB Göttingen
Am Papendiek 14, 37075 Göttingen
neuroth@sub.uni-goettingen.de

Göttingen, 28. Oktober 2011

*Projektgruppe Bildung und Forschung der
Enquete-Kommission Internet und digitale Gesellschaft des Deutschen Bundestages*
Schriftliche Stellungnahme zu ausgewählten Themen der Anhörung am 7. Nov 2011

1. Welche Bedeutung und welche Funktionen haben virtuelle Forschungsumgebungen und die dafür notwendigen Infrastrukturen aktuell für die drei Grundaufgaben der Hochschulen (Forschung, Lehre, Studium)?

Die Allianz Initiative der deutschen Wissenschaftsorganisationen hat in einer Arbeitsgruppe Virtuelle Forschungsumgebungen (VRE¹) wie folgt definiert²:

Eine virtuelle Forschungsumgebung ist eine Arbeitsplattform, die eine kooperative Forschungstätigkeit durch mehrere Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht. Inhaltlich unterstützt sie potentiell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse - während sie technologisch vor allem auf Softwarediensten und Kommunikationsnetzwerken basiert. Virtuelle Forschungsumgebungen sind wesentliche Komponenten moderner Forschungsinfrastrukturen.

(Exzellente) Forschung findet zunehmend globaler, vernetzter und kooperativer statt. Herausragende Beispiele dafür sind die Lebenswissenschaften mit dem DNA-Sequenzierer, einem Verfahren zum automatisierten Ablesen der Sequenzinformation von einem DNA-Molekül: Die Sequenzierungsverfahren der zweiten Generation erlauben heutzutage die Analyse des menschlichen Genoms in ca. acht Tagen. Ein weiteres Beispiel ist die Psycholinguistik mit dem Weltatlas der bedrohten Sprachen, für den große Datenmengen durch die Video- und Tonaufzeichnungen anfallen, auf die weltweit für bestimmte Forschungsfragen zugegriffen wird.

Zusammenfassend kann festgehalten werden, dass Wissenschaft heutzutage und in Zukunft vermehrt in großen und internationalen Forschungsverbänden stattfindet, wobei die Forscher gemeinsam auf verteilte Ressourcen zugreifen. Diese Ressourcen können (fachspezifische) Werkzeuge für die Datenprozessierung z.B. linguistische Analyse großer Textmengen sein, Rechenleistung für z.B. die Simulation von Klimavorhersagen, Server mit großem Speicherbedarf z.B. für die astrophysikalische Kartierung des Weltalls oder auch generische Dienste mit Angeboten für ein vertrauensvolles Login und entsprechenden Abrechnungsmechanismen.

¹ Das Akronym VRE leitet sich aus dem anglo-amerikanischen Raum ab: Virtual Research Environment

² Vgl. http://www.allianzinitiative.de/de/handlungsfelder/virtuelle_forschungsumgebungen/definition/

Wissenschaftler setzen zur Beantwortung ihrer Fragen mehr und mehr digitale Methoden ein und greifen auf digitale Forschungsdaten zurück. Mittlerweile gibt es zum Beispiel für die Klimaforschung, Astrophysik, Teilchenphysik, Medizin und einige textbasierte geisteswissenschaftliche Disziplinen Virtuelle Forschungsumgebungen. Je nach Forschungsfrage stehen dabei zum Teil unterschiedliche Technologien im Vordergrund: Die Klimaforscher, Teilchenphysiker und Astrophysiker müssen riesige Datenmengen prozessieren, daher stehen Rechenleistung und permanenter Speicherbedarf im Vordergrund. Die Geisteswissenschaftler arbeiten mit einer unterschiedlichen Anzahl an digitalen Werkzeugen zum Beispiel Zerlegung von größeren Textmengen in ihre Bestandteile, Analyse und Verlinkung einzelner Begriffe mit digitalen Wörterbüchern, Annotationen etc.

Insgesamt steigt die internationale und nationale Bedeutung von Virtuellen Forschungsumgebungen, Forschungsinfrastrukturen und Forschungsdateninfrastrukturen für die Wissenschaft. Dies spiegelt sich auch in europäischen und internationalen Diskussionen, Initiativen und Förderprogrammen wieder. Aktuell formuliert die Europäische Kommission für das 8. Rahmenprogramm mit Start im Jahr 2013 (Horizon 2020) die Inhalte – Forschungsinfrastrukturen³, Forschungsdateninfrastrukturen und Virtuelle Forschungsumgebungen werden dabei eine zentrale Rolle spielen.⁴

Fazit:

Virtuelle Forschungsumgebungen und die dafür nötigen Infrastrukturen spielen eine zunehmend große Rolle für die Wissenschaft. Nur dadurch wird innovative und exzellente Forschung, die im internationalen Kontext wettbewerbsfähig ist, ermöglicht. Bisher gibt es in Deutschland nur vereinzelte Initiativen und Förderprogramme (BMBF, DFG), die national wenig bis nicht koordiniert sind. Ein Gesamtkonzept hierfür fehlt in Deutschland, auch wenn die Kommission Zukunft der Informationsinfrastruktur im April 2011 ein Gesamtkonzept für die Informationsinfrastruktur in Deutschland⁵ vorgelegt hat, welches bis ca. Mitte 2012 vom Wissenschaftsrat evaluiert und mit Empfehlungen versehen wird. Die Entwicklung von VREs und Infrastrukturen muss aktiv vorangetrieben werden und alle Wissenschaftsdisziplinen erreichen. Darüber hinaus muss sicher gestellt werden, dass neue Berufsfelder (z.B. data curator, data librarian) und Qualifizierungsmaßnahmen bis hinein in die Lehre etabliert werden, um eine neue Generation von Experten und Spezialisten im IKT-gestützten Forschungsprozess zu gewinnen. Dies bedeutet auch, dass diese neuen Themenfelder in die Studiengänge integriert werden müssen bzw. neue Studiengänge etabliert werden müssen, wie es in den USA zum Beispiel zurzeit geschieht.

³ Für den Bereich Forschungsinfrastrukturen hat die EU ein eigenes Förderinstrument ESFRI aufgelegt, bisher werden 44 Forschungsverbände gefördert, vgl. <http://ec.europa.eu/research/esfri/>

⁴ Possible content of Horizon 2020 - the next Framework Programme for Research and Innovation. Thematic workshop on Research Infrastructures (Brussels, 4 July 2011), Report on the workshop, http://ec.europa.eu/research/horizon2020/pdf/workshops/research_infrastructures/summary_report_workshop_on_4_july_2011.pdf#view=fit&pagemode=none

⁵ <http://www.leibniz-gemeinschaft.de/?nid=kiikom&nidap=&print=0>

2. Wie ist mit Blick auf die virtuellen Forschungsumgebungen der aktuelle Stand der technischen Ausstattung, welche Standards und Nutzungsarten in den verschiedenen Fachbereichen gibt es und welche Voraussetzungen sind bei (internationalen) Kooperationen notwendig?

Allgemein kann gesagt werden, dass Virtuelle Forschungsumgebungen - unabhängig vom Fachbereich - darauf basieren, dass Ressourcen gemeinsam genutzt werden⁶. Ressourcen können dabei sein:

- gemeinsamer, standortunabhängiger Zugriff auf **Daten und Forschungsdaten**
- Zugriff auf **Speicherplatz**, dauerhaft und/oder temporär
- Zugriff auf **Rechenleistung**
- Gemeinsamer Zugriff auf **Werkzeuge und Tools**, die nicht mehr auf jedem Arbeitsplatzrechner installiert sind, sondern bei Bedarf dazugeschaltet werden können
- Nutzung einer (national/international) koordinierten **Basis-Infrastruktur**, die alle Aspekte eines Forschungsprozesses unterstützt

Es gibt zurzeit in Deutschland bisher nur wenige Virtuelle Forschungsumgebungen (VRE), die im Sinne der o.g. Definition operieren. Im Rahmen der D-Grid Initiative wurden mit mehr als 100 Mio Euro eine Reihe von Vorhaben gefördert wie zum Beispiel Infrastruktur⁷ kommerzielle und akademische Projekte. Die u.a. im Rahmen der D-Grid Initiative⁸ finanzierten akademischen VREs (Teilchenphysik, Klimaforschung, Medizin, Geisteswissenschaften, Astrophysik) können die Basis-Technologien der Grid-Initiative nutzen (z.B. Authentifizierung, Monitoring, Accounting, Middleware etc.).

Im Rahmen des BMBF finanzierten Projektes WissGrid- Grid für die Wissenschaft⁹ bauen diese fünf Virtuellen Forschungsumgebungen ihre technische Kooperation weiter aus, zum Beispiel im Bereich der Langzeitarchivierung von Forschungsdaten. Das Projekt WissGrid hat gezeigt, dass auch bei großen disziplinären Unterschieden und Anforderungen, gemeinsame Technologien erhebliche Synergieeffekte bringen und zumindest generische Technologie-Komponenten gemeinsam genutzt werden können. Damit hat sich klar gezeigt, dass auch dieser von der EU vorgezeichnete Weg zum Aufbau fachübergreifender technologischer Infrastrukturen der einzige erfolgversprechende Ansatz ist, um Virtuelle Forschungsumgebungen in möglichst vielen Fachbereichen langfristig zu befördern. Die IT-Expertise ist in den verschiedenen Fachdisziplinen jeweils sehr unterschiedlich ausgeprägt und nur ein gemeinsames Verständnis über eine technologische Basis-Infrastruktur und deren dauerhafte Bereitstellung sorgt in der Forschung und Lehre für die nötigen Impulse, neue Forschungsfragen zu generieren und alte Forschungsfragen mit neuen digitalen Methoden zu bearbeiten. Folgende technologische Anforderungen stehen dabei im Vordergrund, die jeweils unterschiedliche Nutzungsarten bedienen:

⁶ Eine Übersicht über „technische“ Ressourcen, die im Rahmen der D-Grid Initiative gefördert wurden gibt <http://www.d-grid-ggmbh.de/index.php?id=10>

⁷ Vgl. <http://www.d-grid-ggmbh.de/index.php?id=71>

⁸ <http://www.d-grid.de/>

⁹ <http://www.wissgrid.de/>

- **Umgang mit großen digitalen Datenmengen** (Generierung z.B. bei Großinstrumenten, Prozessierung, Verarbeitung etc.)¹⁰
- **Dauerhafte Bereitstellung und Nachnutzung von Forschungsdaten**, Aufbau von (national, international) koordinierten Forschungsdateninfrastrukturen und Forschungs-Datenzentren¹¹
- **Bereitstellung von (temporärer) Rechenleistung** „on demand“¹²
- Entwicklung **generischer und fachspezifischer Werkzeuge**, die sowohl disziplin-übergreifend als auch in einzelnen Disziplinen genutzt werden können
- (Internationale) Entwicklung von (fachspezifischen, offenen) **Standards** wie zum Beispiel im Bereich von Schnittstellen, Metadaten, Datei-Formate etc.

Um international kooperieren zu können, müssen die Virtuellen Forschungsumgebungen internationale Entwicklungen und Standards berücksichtigen. In der heutigen global vernetzten Welt sind die Fachwissenschaftler mehr und mehr international vernetzt und sie möchten sich unabhängig von vorgegeben Technologien austauschen, gemeinsam fachspezifische Werkzeuge entwickeln sowie nutzen, Daten und Erkenntnisse austauschen und dabei bei Bedarf schnell und unkompliziert Spitzentechnologie wie Rechenleistung und Speicherplatz abrufen. Dafür ist es notwendig sowohl in Deutschland, aber auch international eingebettet in dortige Entwicklungen, forschungsnahe Infrastrukturen aufzubauen und nachhaltig bereit zu stellen, die die einzelnen Fachbereiche bei ihren jeweils aktuellen Forschungsfragen maximal unterstützt. Die Situation in der Forschung und Lehre ist vergleichbar mit den Infrastrukturen, die im Bereich der Mobilität entwickelt wurden bzw. werden: Je schneller es möglich ist, von Ort A nach Ort B zu reisen, Informationen und Erkenntnisse auszutauschen und je einfacher dabei die standardisierten Formalitäten und Abwicklungen sind, desto schneller schreitet der (mobile) Fortschritt voran. Dabei ist es unerlässlich, eine zum Beispiel gemeinsame Infrastruktur (z.B. im Flugverkehr) aufzubauen, sich auf Standards in der Abwicklung (Pass-Kontrolle, Flugsicherheit etc.) zu verständigen, Richtlinien zu definieren, die technische Basis-Infrastruktur (Flughäfen) und spezifische Infrastruktur-Komponenten (Check-In, Sicherheitskontrollen) zu entwickeln sowie permanent weiterzuentwickeln, zu optimieren und auch dauerhaft bereit zu stellen.

¹⁰ Vgl. z.B. das Tier-Konzept bei CERN für die Datenspeicherung, <http://lcg.web.cern.ch/lcg/public/tiers.htm>

¹¹ World Data Center System (wovon es bereits einige in Deutschland gibt), <http://www.ukssdc.ac.uk/wdcmmain/>

¹² Vgl. HPC Computing, http://en.wikipedia.org/wiki/High-performance_computing und in Deutschland die Gauß-Allianz, <http://gauss-allianz.net/de/ueber-die-gauss-allianz.html>

3. Wie müsste Ihrer Einschätzung nach ein Ausbau virtueller Forschungsumgebungen systematisch und kontinuierlich vorangetrieben und finanziert werden?

Um Deutschland (weiterhin) im Bereich Forschung und Lehre wettbewerbsfähig zu halten, braucht es folgende Aktivitäten:

- **Bewusstsein schaffen** für die neue Ära von Wissenschaft: in der Politik, Wissenschaft, Gesellschaft und bei Förderern. Die neue Forschung ist Technologie-gestützt und basiert auf immensen, immer größer werdenden, digitalen Datenmengen¹³.
- Einrichtung einer **nationalen Expertengruppe**, die Disziplin-spezifisch und Disziplin-übergreifend die Aktivitäten koordiniert, Anforderungen definiert, Umsetzungen begleitet und Empfehlungen formuliert für die Weiterentwicklung und internationale Konkurrenz-sowie Anpassungsfähigkeit.
- Etablierung **geeigneter Geschäftsmodelle** inklusive Organisations- und Finanzierungsmodelle, die sowohl die Institutionen-übergreifende Kooperation erlauben als auch den föderalen Strukturen (Bund-Land-Problematik) Rechnung tragen.
- (Neu)Definition von **Rollen, Funktionen und Verantwortlichkeiten** aller beteiligten Stakeholder (Bibliotheken, Rechenzentren, Wissenschaftsdisziplinen, Datenzentren etc.), um alle Aspekte von der Datenerhebung/-generierung, Datenweiterverarbeitung/-prozessierung bis hin zur Datenpublikation abzudecken.
- Entwicklung **geeigneter Qualifizierungsmaßnahmen** und **neuer Berufszweige**, die als eigenständige Studiengänge und in Form von Schlüsselkompetenzen in die Lehre und in das Studium integriert werden.
- Unterstützung bei der Entwicklung von **Standards und Werkzeuge** für die Datenanalyse, Datenanreicherung und Datenpublikation.

¹³ Riding the wave - How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission, October 2010, <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>