

## BERICHT ÜBER DEN WORKSHOP „OFFENER ZUGANG ZU DATEN – EINE HERAUSFORDERUNG“ IM RAHMEN DER OPEN-ACCESS-TAGE 2008 AM 10.10.2008 IN BERLIN<sup>92</sup>

von Heinz Pampel

### Inhalt und Diskussion

Der Workshop „Offener Zugang zu Daten – eine Herausforderung“<sup>93</sup> begann mit einem Impulsvortrag „Offener Zugang zu Forschungsdaten - eine Herausforderung“ von Dr. Hans Pfeiffenberger.<sup>94</sup> In diesem beschrieb der Referent, ausgehend von der finanziellen Dimension, die Relevanz des Themas: Die Erhebung von Forschungsdaten sei häufig sehr kostenintensiv. Am Beispiel der Geowissenschaften nannte er neben den Personalkosten die Unterhaltung von Großgeräten als einen signifikanten Kostenfaktor. Im Rahmen einer „guten wissenschaftlichen Praxis“<sup>95</sup> stehe neben der Verantwortung gegenüber dem Steuerzahler die Effektivität der Forschung im Zentrum. Forschungsdaten nachnutzbar zu machen, sei eine gesellschaftliche Verantwortung der Wissenschaft. Dank einer offenen Zugänglichkeit der Forschungsdaten werde der Fortschritt der Wissenschaft begünstigt, da, neben der Nachprüfbarkeit der publizierten Ergebnisse, die Forschung auf Basis bereits veröffentlichter Daten ermöglicht werde.

Weiter machte Pfeiffenberger deutlich, dass sich der Wissenschaft mit dem Fortschritt der Informations- und Kommunikationstechnologie neue Möglichkeiten eröffnet haben, die im angloamerikanischen Raum mit den Stichwörtern *E-Science* und *Cyberinfrastructure* beschrieben werden. Die stetig steigende Menge an Forschungsdaten und die in vielen Disziplinen bereits etablierte kollaborative Forschung fordern neue Strukturen – organisatorischer und technischer Art – im Umgang mit den Forschungsdaten.

Diese Notwendigkeit, so Pfeiffenberger, zeige sich auch in der wissenschaftspolitischen Diskussion. Die im Juni 2008 von den *European Heads Of Research Councils* (EUROHORCs) und der *European Science Foundation* (ESF) veröffentlichte Vision des europäischen Forschungsraums „The EUROHORCs

---

<sup>92</sup> Die Open-Access-Tage 2008 fanden vom 9. bis 10.10.2008 auf dem Campus der Freien Universität Berlin statt und wurden in Kooperation mit der Helmholtz-Gemeinschaft und der Max-Planck-Gesellschaft veranstaltet. Die Dokumentation der Veranstaltung ist auf der Informationsplattform [open-access.net](http://open-access.net) unter folgender URL auffindbar:

[http://open-access.net/de/aktivaeten/openacesstage/archiv/openacesstage\\_2008/#c1680](http://open-access.net/de/aktivaeten/openacesstage/archiv/openacesstage_2008/#c1680)

Ein ausführlicher Tagungsbericht findet sich in: Vock, Rubina: Freier Zugang zu Wissen. Open-Access-Tage 2008. In: H-Soz-u-Kult, 05.12.2008. URL: <http://hsozkult.geschichte.hu-berlin.de/tagungsberichte/id=2400>

<sup>93</sup> Der Workshop „Offener Zugang zu Daten – eine Herausforderung“ wurde von Dr. Jens Klump (Deutsches GeoForschungs-Zentrum GFZ, <http://www.gfz-potsdam.de>) und Dr. Hans Pfeiffenberger (Alfred-Wegener-Institut für Polar- und Meeresforschung – AWI, <http://www.awi.de>) organisiert. Beide sind in Einrichtungen der Helmholtz-Gemeinschaft (<http://www.helmholtz.de>) tätig und beschäftigen sich seit mehreren Jahren mit den Herausforderungen rund um den Open Access zu Forschungsdaten. Klump ist wissenschaftlicher Mitarbeiter am Daten- und Rechenzentrum des GFZ. Er ist u.a. für die Scientific Drilling Database (SDDB, <http://www.icdp-online.org/contenido/lakedb>) des GFZ zuständig und in nestor, dem deutsche Kompetenznetzwerk zur digitalen Langzeitarchivierung (<http://www.langzeitarchivierung.de>), sowie in dem von der Deutschen Forschungsgemeinschaft (DFG, <http://www.dfg.de>) geförderten Projekt zur Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten (STD-DOI, <http://www.std-doi.de>) involviert. Pfeiffenberger ist stellvertretender Leiter des Rechenzentrums des AWI. In einer Vielzahl von Projekten beschäftigt er sich u.a mit den Themen Repositorien, Grid-Computing und Informations- und Kommunikationstechnologien für wissenschaftliche Institutionen. Er ist Sprecher des Arbeitskreises Open Access in der Helmholtz-Gemeinschaft (<http://oa.helmholtz.de>) und in der Arbeitsgruppe Forschungsprimärdaten der Allianz der deutschen Wissenschaftsorganisationen aktiv. Darüber hinaus ist er Herausgeber der Open-Access-Zeitschrift *Earth System Science Data* (ESSD, <http://www.earth-system-science-data.net>).

<sup>94</sup> Klump, Jens; Pfeiffenberger, Hans: Offener Zugang zu Forschungsdaten - eine Herausforderung. Open-Access-Tage 2008. Berlin, 2008. URL: <http://open-access.net/fileadmin/OAT/Pfeiffenberger-OAT-2008-10-10.pdf>

<sup>95</sup> Deutsche Forschungsgemeinschaft: Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission Selbstkontrolle in der Wissenschaft. Weinheim, 1998. URL: [http://www.dfg.de/aktuelles\\_presse/re-den\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_0198.pdf](http://www.dfg.de/aktuelles_presse/re-den_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf)

and ESF Vision on a Globally Competitive ERA and their Road Map for Actions to Help Build it“<sup>96</sup> und die im selben Monat verabschiedete Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen widmen sich dem Umgang mit Forschungsdaten. Während in der EUROHORCs/ ESF-Vision Open Access zu Forschungsdaten gefordert werde – „Open access to the output of publicly funded research and permanent access to primary quality assured research data“ – formulieren die deutschen Wissenschaftsorganisationen einen „dringenden Handlungsbedarf hinsichtlich der systematischen Sicherung, Archivierung und Bereitstellung dieser Daten für die Nachnutzung durch Dritte.“<sup>97</sup>

Anhand des Internationalen Polarjahres 2007-2008<sup>98</sup> und anderer Aktivitäten in den Geowissenschaften machte Pfeiffenberger anschließend auf die praktischen Herausforderungen des Managements von Forschungsdaten aufmerksam. Dabei beschrieb er unter anderem die Notwendigkeit von Richtlinien, Repositorien, qualifizierten Akteuren, Strategien der Langzeitarchivierung und disziplinspezifischen Lösungsansätzen.

Zum Schluss stellte der Referent die Open-Access-Zeitschrift *Earth System Science Data* (ESSD) vor, deren Herausgeber er ist. Diese widme sich der Beschreibung von frei zugänglichen Forschungsdaten. Dabei, so Pfeiffenberger, werde im Rahmen eines Peer Review die Begutachtung der Forschungsdaten sowie die Zitierbarkeit der Daten gewährt. Dank dieser Strategie werden entscheidende Anreize zur frei zugänglichen Veröffentlichung von Forschungsdaten geschaffen.

Anschließend widmete sich Dr. Jens Klump den Voraussetzungen des Open Access zu Forschungsdaten. Dabei benannte er sechs Aspekte, die den systematischen Umgang – im Rahmen von Open Access – mit Forschungsdaten aus Sicht der aktuellen Diskussion bestimmten: „Vertrauenswürdige Daten (Qualitätssicherung), Langzeitverfügbarkeit der Daten (Archive), Nachweis des Bestands (Kataloge), stabile Links (persistente Identifikatoren), klare Regeln für die Nutzung (Lizenzen), organisatorischer Rahmen (Workflows).“<sup>99</sup> Diese Aspekte wurden mit den rund 40 Teilnehmerinnen und Teilnehmern des Workshops diskutiert. Dabei wurden die unterschiedlichen Herangehensweisen der Teilnehmenden an die Problematik deutlich. Je nach fachlichem Hintergrund der Teilnehmerinnen und Teilnehmer standen verschiedene Vorstellungen von Forschungsdaten im Raum, was sich zum Beispiel in der Verwendung unterschiedlicher Terminologien (Primär- und Sekundärdaten; Daten; Forschungsdaten) äußerte.<sup>100</sup>

Diese disziplinspezifischen Perspektiven schlugen sich auch bei der Diskussion um das Kernthema Open Access nieder. Während Pfeiffenberger, aufgrund seiner positiven Erfahrungen in den Geowissenschaften, zu offen zugänglichen Forschungsdaten warb, äußerten sich andere Teilnehmerinnen und Teilnehmer, mit Blick auf disziplinspezifische Herausforderungen, skeptisch. Konsens herrschte darüber, dass Forschungsdaten, welche Grundlage einer wissenschaftlichen Publikation sind, frei zugänglich sein sollten. Eine Forderung, die auch von naturwissenschaftlichen Verlagen, im Rahmen der „Brüssel Declaration“, mitgetragen wird.

---

<sup>96</sup> European Heads Of Research Councils; European Science Foundation (2008): The EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions to Help Build it.

URL:[http://eurohorcs.drift.senselogic.se/download/18.45b270a411a9ed8e12780003647/EUROHORCs\\_ESF\\_ERA\\_RoadMap.pdf](http://eurohorcs.drift.senselogic.se/download/18.45b270a411a9ed8e12780003647/EUROHORCs_ESF_ERA_RoadMap.pdf)

<sup>97</sup> Allianz der deutschen Wissenschaftsorganisationen: Schwerpunktinitiative Digitale Information der Allianz-Partnerorganisationen. 2008. URL: [http://www.dfg.de/aktuelles\\_presse/das\\_neueste/download/pm\\_allianz\\_digitale\\_information\\_details\\_080612.pdf](http://www.dfg.de/aktuelles_presse/das_neueste/download/pm_allianz_digitale_information_details_080612.pdf)

<sup>98</sup> <http://www.polarjahr.de>

<sup>99</sup> Pfeiffenberger, Hans; Klump, Jens: Offener Zugang zu Daten. Open-Access-Tage 2008. Berlin, 2008. URL: [http://open-access.net/fileadmin/OAT/Workshop\\_Daten.pdf](http://open-access.net/fileadmin/OAT/Workshop_Daten.pdf)

<sup>100</sup> Zur Problematik der Begriffe Primär- und Sekundärdaten siehe: Klump, Jens: Wissenschaftliche Primärdaten. Version 1.5. In: nestor-Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, 2008. URL: [http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor\\_handbuch\\_artikel\\_275.pdf](http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_275.pdf). In Deutschland scheint sich der Begriff Forschungsdaten zu etablieren, der insbesondere die interdisziplinäre Diskussion vereinfacht.

In der Diskussion zum Themenfeld „Vertrauenswürdige Daten (Qualitätssicherung)“ stand die Frage nach Kriterien zur Qualitätssicherung im Mittelpunkt. Klump machte deutlich, dass die Nachnutzung eine hohe Qualität der Forschungsdaten voraussetze. Hier, so das Fazit aus der Diskussion, sind Erhebende und Nachnutzende gefordert, sich auf disziplinspezifische Kriterien der Qualitätssicherung festzulegen. Bei der „Langzeitverfügbarkeit der Daten (Archive)“ scheint sich ein disziplinübergreifendes Vorgehen zu bewähren. Klump wies auf die Empfehlung Nr. 7 zur Sicherung guter wissenschaftlicher Praxis hin, welche in Deutschland Standard sei:

„Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“<sup>101</sup>

In diesem Kontext machte Klump auf die völlig unterschiedlichen Auffassungen des Begriffes „Langzeit“ aufmerksam. Während Bibliotheken, Archive und Museen (je nach Sammelauftrag) mehrheitlich die Dauerhaftigkeit des Begriffs über viele Jahrhunderte hinweg betonen, werde in der Informatik unter Langzeit ein deutlich kürzerer Zeitraum verstanden. In der wissenschaftlichen Praxis verkürze sich die Verfügbarkeit der jeweiligen Forschungsdaten häufig auf den jeweiligen Projektzeitraum. Eine mögliche Lösung, Forschungsdaten systematisch zu managen, sei in der Gestaltung des „organisatorischen Rahmens (Workflows)“ zu finden. Klump beschrieb das an der Monash University entwickelte Modell des *Data Curation Continuum*.<sup>102</sup> Dieses beschreibe, gegliedert in drei Stationen (*Privat Research Domain, Shared Research Domain, Public Domain*) den Weg digitaler Objekte aus der geschlossenen Arbeitsplattform der Wissenschaftlerin oder des Wissenschaftlers in das frei zugängliche Repositorium der jeweiligen Institution. Das Modell stieß unter den Anwesenden auf Interesse. In Fragen und Kommentaren wurden insbesondere die Anforderungen an die einzelnen Schnittstellen zwischen den Stationen deutlich, in welchen Migrationen stattfinden. Dabei wurde deutlich, dass diese Schnittstellen neben einer technischen auch eine organisatorische Dimension haben. Mehreren Teilnehmerinnen und Teilnehmer wiesen in diesem Kontext auf fehlende Konzepte und Werkzeuge hin.

Sobald Forschungsdaten nachnutzbar sind, ergibt sich das Problem „stabiler Links (persistente Identifikatoren)“. Klump berichtete über die Erfahrungen des von der DFG geförderten Projekts zur Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten (STD-DOI). In diesem sei die Adressierung von Forschungsdaten mit persistenten Identifikatoren erfolgreich erprobt worden. Einhergehend mit der stabilen Identifikation der Forschungsdaten gelten die von Pfeiffenberger im Rahmen der Zeitschrift *Earth System Science Data* gemachten Erfahrungen zur Zitation der Forschungsdaten als entscheidender Anreiz. Aufbauend auf der Voraussetzung der persistenten Adressierung gelte es, so Klump, den „Nachweis des Bestands (Kataloge)“ zu garantieren. Die Diskussion zu diesem Thema berührte die Anforderungen an Metadaten und die dafür in vielen Disziplinen schon etablierten Standards sowie die Weiterentwicklung von Suchtechnologien. Klump nahm hier die Sichtweise des praktizierenden Wissenschaftlers ein und forderte zentrale Zugänge.

Weiterhin müssen nach Klumps Ansicht „klare Regeln für die Nutzung (Lizenzen)“ von Forschungsdaten formuliert werden. In der Diskussion zu diesem Punkt wurde das Spannungsfeld zwischen Open Access, wirtschaftlichen Interessen in einigen Disziplinen (z.B. im Rahmen von Patenten) und soziologischen Aspekten deutlich. Auch hier, so das Fazit der Diskussion, gilt es disziplinspezifische Antworten zu finden.

---

<sup>101</sup> [http://www.dfg.de/aktuelles\\_presse/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_0198.pdf](http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf)

<sup>102</sup> Eine Beschreibung des Modells findet sich in folgenden Publikationen: Treloar, Andrew; Groenewegen, David; Harboe-Ree, Cathrine: The Data Curation Continuum Managing Data Objects in Institutional Repositories. In: D-Lib Magazine 13 (2007) 9/10. DOI:10.1045/september2007-treloar; Treloar, Andrew; Harboe-Ree, Cathrine: Data management and the curation continuum. How the Monash experience is informing repository relationships. VALA2008 14th Biennial Conference. Melbourne, 2008. URL: [http://www.valaconf.org.au/vala2008/papers2008/111\\_Treloar\\_Final.pdf](http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf)

## Fazit

Das starke Interesse der Teilnehmer und Teilnehmerinnen aus Bibliotheken, Rechenzentren und der Wissenschaft zeigte die Aktualität des Themas. Deutlich wurde, dass die im anglo-amerikanischen Raum geführte Diskussion und die bereits entwickelten Konzepte in Deutschland größtenteils unbekannt sind.<sup>103</sup> Die Referenten zeigten, dass der freie Zugang zu Forschungsdaten einen systematischen Umgang mit diesen voraussetzt.

Die Diskussion während des Workshops bewegte sich häufig auf einer sehr abstrakten Ebene. In vielen Disziplinen, so scheint es, fehlen die Erfahrungen mit frei zugänglichen Forschungsdaten. Während allgemein der freie Zugang zu Forschungsdaten gefordert wird, sind es häufig detaillierte disziplinspezifische Herausforderungen, die eine interdisziplinäre Diskussion erschweren. Die von Pfeifferberger und Klump vorgestellte Strategie der Publikation von Forschungsdaten ähnlich einer traditionellen Text-Publikation scheint aktuell die wohl konsensfähigste Lösung zu sein.

---

<sup>103</sup> Eine umfassende Einführung findet sich z.B. bei Lyon, Liz: Dealing with Data. Roles, Rights, Responsibilities and Relationships. Consultancy Report. 2007.

URL: [http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing\\_with\\_data\\_report-final.pdf](http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf)