



Symposium Forschungsdaten-Infrastruktur (FDI 2013) 22. Januar 2013, GeoForschungsZentrum Potsdam (GFZ)

Gemeinsam organisiert von den DFG-Projekten
Radieschen, re3data.org, KomFor, EWIG und BoKeLa

Symposium-Bericht

Die Flut an digitalen Daten in Wissenschaft und Forschung wächst rasant. Die dauerhafte Speicherung dieser Daten für zukünftige Generationen von Forscherinnen und Forschern stellt das gesamte Wissenschaftssystem vor Herausforderungen. Jüngst hat die EU-Kommission Empfehlungen an ihre Mitgliedsstaaten zur dauerhaften Zugänglichkeit von Forschungsdaten verabschiedet. Doch noch sind viele Fragen ungelöst. So müssen Aspekte der Finanzierung, Organisation und Technologie der zu schaffenden Forschungsdaten-Infrastrukturen, sowie deren rechtliche und politische Rahmenbedingungen geklärt werden. Diese Themenfelder wurden im Rahmen eines gemeinsam organisierten Symposiums der DFG-Projekte Radieschen, re3data.org, KomFor, EWIG und BoKeLa diskutiert.

Im Hauptprogramm des Symposiums wurden, mit dem Datenlebenszyklus als Leitlinie, die jeweiligen Herausforderungen und neuesten Erkenntnisse durch Vorträge und daran anschließende Diskussionen erörtert. In begleitenden Workshops wurden zentrale Themenfelder von Expertinnen und Experten bearbeitet und deren Ergebnisse anschließend dem Publikum vorgestellt.

Das Symposium versammelte die Experten-Community zu Forschungsdaten aus dem deutschsprachigen Raum. Kern der Veranstaltung bildeten Vorträge zu verschiedenen Aspekten der Forschungsdaten-Infrastrukturen, sowie Präsentationen laufender Forschungsprojekte. Darüber hinaus wurden zentrale Herausforderungen der Weiterentwicklung von Forschungsdaten-Infrastrukturen in Workshops diskutiert. Desweiteren bot das Symposium eine Plattform zum Gedankenaustausch und zum Netzwerken.

Das Symposium fand am 22. Januar 2013 am Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum GFZ statt. Eröffnet wurde die Veranstaltung durch den Vorstand des GeoForschungsZentrums **Prof. Dr. Dr. h.c. Reinhard F.J. Hüttl** und durch ein Geleitwort von **Dr. Stefan Winkler-Nees (DFG)**. Das Vortragsprogramm begann mit einer Keynote von Dr. Torsten Reimer vom Joint Information Systems Committee (JISC).



Abb. 1: Begrüßung durch den Vorstand des GFZ – Prof. Dr. Dr. h.c. Reinhard F.J.Hüttl

Gefördert von:

DFG Deutsche
Forschungsgemeinschaft

Keynote

Dr. Thorsten Reimer (JISC) beschrieb in seiner Keynote die Herausforderungen im Aufbau von Forschungsdaten-Infrastrukturen und erläuterte die Vorteile der dauerhaften Zugänglichkeit von Daten für die verschiedenen Stakeholder. Nach einer kurzen Vorstellung von JISC erläuterte Dr. Reimer Visionen für den Entwicklungsstatus von Forschungsdaten - Infrastrukturen mit 10 Jahresperspektive.

JISC verfolgt in Großbritannien einen 4-Phasenplan für die Entwicklung von Forschungsdaten-Infrastrukturen und dem Management von Forschungsdaten, an dem sich auch das zugehörige vierstufige Förderprogramm orientiert. Momentan befindet sich das Programm in der Konsolidierungsphase, in der es darum geht, gefundene Lösungen zu verfestigen und institutionelle Kapazitäten weiter auszubauen. In diesem Kontext beschrieb er zwölf damit verbundene Herausforderungen und verschiedene Ansätze, diesen Herausforderungen zu begegnen.

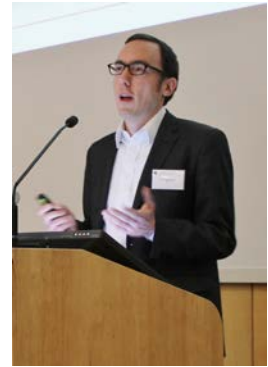


Abb. 2: Keynote-Sprecher Dr. Thorsten Reimer, JISC

Präsentation der DFG-Projekte Radieschen, re3data.org, KomFor, EWIG und BoKeLa

Frank Scholze, KIT, stellte das Projekt re3data.org – Registry of Research Data Repositories vor. Das Projekt hat das Ziel Daten-Repositoryn in einem web-basierten Verzeichnis zu erschließen und so eine Orientierung über bestehende Datensammlungen zu bieten. Zentrales Anliegen von re3data.org ist es, Wissenschaftlern eine Orientierung in der heterogenen Landschaft der Forschungsdaten-Repositoryn zu geben. Dabei wird sowohl die Rolle der Erhebenden, die z. B. durch Förderorganisationen oder wissenschaftliche Zeitschriften aufgefordert sind Daten zugänglich machen, als auch die Rolle der Nutzenden, z. B. Forschende, die Daten von Dritten nachnutzen möchten, fokussiert. Weiter soll infrastrukturellen Dienstleistern, wie Datenzentren, Rechenzentren und Bibliotheken, eine Übersicht über die Landschaft der Forschungsdaten-Repositoryn gegeben werden. Darüber hinaus kann der Dienst von Förderorganisationen genutzt werden: Fordern diese Mittelempfänger auf, gewonnene Forschungsdaten offen zugänglich zu machen, kann eine Recherche in re3data.org helfen, potenzielle Repositoryn zur Speicherung der Forschungsdaten zu finden. Weiterführende Informationen finden sich auf der Projekt-Website (<http://www.re3data.org>).



Abb. 3: Präsentation „re3data.org“, Frank Scholze, KIT

Tim Hasler, ZIB, präsentierte das DFG-Projekt EWIG - Entwicklung von Workflowkomponenten für die Langzeitarchivierung von Forschungsdaten in den Geowissenschaften. EWIG zielt auf Probleme speziell bei der Datenübergabe an Langzeitarchive. Dazu werden drei Themen beleuchtet. Zum einen werden auf institutioneller Ebene bei den Projektpartnern GFZ und Institut für Meteorologie der FU Berlin Policies entwickelt, die die Grundlage für eine genaue Übergabedefinition zwischen Datenlieferant und Archiv / Repository sind. Am Zuse-Institut Berlin liegt der Fokus auf der technischen Qualitätssicherung gelieferter Daten. Der dritte Schwerpunkt der Arbeiten soll das Thema Forschungsdatenmanagement bereits in der fachwissenschaftlichen Ausbildung etablieren. Dazu werden in EWIG Konzepte für Lehr- und Weiterbildungsveranstaltungen entworfen. <http://ewig.gfz-potsdam.de/>.

Dr. Jochen Klar, AIP, präsentierte das DFG-Projekt „Rahmenbedingungen einer disziplin-übergreifenden Forschungsdateninfrastruktur“, kurz Radieschen. Ziel des Projekts ist die Erstellung einer Roadmap mit Handlungsempfehlungen für eine disziplinübergreifende Infrastruktur für Forschungsdaten in Deutschland.

Sie identifiziert und behandelt die Anforderungen an generische Komponenten einer Infrastruktur und die Vernetzung mit disziplinspezifischen Komponenten. Die Analyse basiert auf einer Bestandsaufnahme mit bestehenden und neuen Projekten sowie Maßnahmen zum Community Building. Zentrale Dimensionen der Analyse sind Technik, Organisation und Kosten sowie die Untersuchung von Querschnittsthemen. Das Projektkonsortium ist ebenso disziplin- und organisationsübergreifend wie die Analyse. Weiterführende Informationen finden sich auf der Projekt-Website unter <http://www.forschungsdaten.org/>.

Dr. Michael Diepenbroek, MARUM, präsentierte das DFG-Projekt KomFor – Kompetenzzentrum für Forschungsdaten aus Erde und Umwelt (<http://www.komfor.net>). Das Kompetenzzentrum ist als Bindeglied zwischen Forschungseinrichtungen, Verlagen, Bibliotheken und einem bestehenden Archivnetzwerk für Daten aus Erd- und Umweltforschung geplant. Allgemeines Ziel ist die nachhaltige Verbesserung von Datenverfügbarkeit und –qualität. Konkret sollen nachhaltige und verlässliche Wege zur Publikation wissenschaftlicher Daten geschaffen werden die den Qualitätsstandards wissenschaftlichen Publizierens entsprechen. Das Kompetenzzentrum soll hierfür wissenschaftliche Projekte, Institute, Forschergruppen aber auch einzelne Wissenschaftler in allen Fragen des Datenmanagements begleiten – von der Planungsphase über die Datenerhebung, Qualitätssicherung, Registrierung und Langzeitarchivierung bis zur Publikation der Daten. Weiterführende Informationen zum Projekt auf <http://www.komfor.net/>.



Abb. 4: Präsentation „KomFor“, Dr. Michael Diepenbroek, MARUM

Dr. Hans-Joachim Wallrabe-Adams, MARUM, erläuterte in seinem Vortrag die Ziele und aktuellen Ergebnisse des Projekts BoKeLa - Aufbau des Dateninformationssystems für das GESEP Kern- und Probenlager zur Erfassung und Verwaltung von Bohrkernen und Nachweis der Bestände in einem Internetportal (<http://www.gesep.de>). Das Projekt BoKeLa wurde vom German Scientific Earth Probing Consortium (GESEP e.V.) initiiert. Dem Konsortium gehören 15 geowissenschaftliche Institute mit einer breiten Expertise im Bereich Forschungsbohrungen an Land, im Meer und im Eis an. BoKeLa wird von drei Projektpartnern getragen: der Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), dem Deutschen GeoForschungszentrum (GFZ) und dem Zentrum für Marine Umweltwissenschaften (MARUM). Ausgangspunkt der Entwicklung eines Bohrkern- und Probenverwaltungssystems ist das Drilling Information System (DIS) das bereits seit Jahren in diesem Bereich eingesetzt wird, aber für die hier geplante Nutzung umfangreicher Anpassungsarbeiten bedarf. Dieses System unterstützt das Datenmanagement bereits an der Bohrung, dann im Labor und dient später der Verwaltung im Kernlager. Alle Kerne und Proben werden mit einem eindeutigen Identifikator versehen, der International Geo Sample Number (IGSN). Damit wird es möglich sein Kerne und Proben dauerhaft identifizierbar und in der Fachliteratur zitierbar zu machen, sowie wissenschaftliche Daten direkt Proben zuzuordnen. Über ein Webportal werden die Kernlager-Informationen nach dem Open Access Prinzip zugänglich gemacht und mit Datenzentren (z.B. PANGAEA) und Publikationen verknüpft.

Vorträge

Die Vorträge des Hauptprogramms spiegelten die einzelnen Stationen im Datenlebenszyklus wieder. Abb. 5 zeigt das Domänenmodell nach Treloar mit integrierter dauerhafter Domäne.

Als ein Beispiel für die Private Domäne erläuterte **Prof. Frederik Tilman (GFZ)** den Datenzyklus in der Wissenschaft aus der Perspektive der Forscher anhand seines Forschungsgebiets Seismologie. Als Herausforderung beschrieb er das Auffinden von früher datierten Datensets, die Verwaltung der Daten, das Vermeiden von Überschreibungen, sowie die Nicht-Reproduzierbarkeit der Daten aufgrund fehlender oder veralteter Software. Eingesetzte Software-Lösungen sind zumeist ‚low-tech solutions‘.

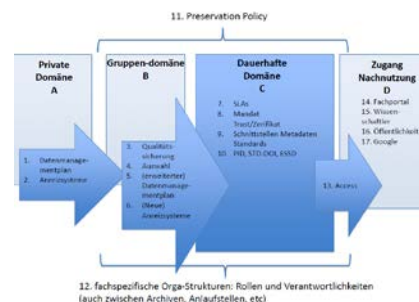


Abb. 5: Domänen-Modell
(abgeleitet von Treloar, A. &
Harboe-Ree, 2008)

Eine Herausforderung sei es, laut Prof. Tilman, die Wissenschaftler dazu zu bewegen, komplexere, aber besser geeignete Lösungen zum informellen Datenaustausch zu verwenden. Prof. Tilman zeigte persönliche, soziale und technologische Hinderungsgründe auf, die für viele Wissenschaftler dem freien Datenaustausch entgegenstehen. Als mögliche Anreize, um dieses Problem zu beheben, nannte er die gesteigerte Anerkennung von Datenpublikationen als wissenschaftliche Leistungen, die Einführung von Datenpublikationen als Voraussetzung für Publikationen generell, sowie mehrere Anreize, die auf das persönliche Umfeld der Forscher abzielen. Ein wichtiger Punkt sei auch die Nachvollziehbarkeit des zeitlichen Ablaufs sowohl für die Erstellung der Daten als auch für den Einsatz der Software. Er schloss seinen Vortrag mit dem Wunsch nach einfach verfügbaren und leicht auffindbaren Datensets, der Forderung nach wiederverwendbaren und rekonstruierbaren Datenprodukten und der Hoffnung auf bessere Anerkennung für die offene Weitergabe und das Publizieren von Daten.

Dr. Peter Bartelheimer (Universität Göttingen) betrachtete in seinem Vortrag die Datenbereitstellung und Virtuelle Forschungsumgebungen (VFU) als zwei Komponenten der IT-Forschungsinfrastruktur. Anhand des Projekts „Virtuelle Forschungsumgebung für die sozio-ökonomische Berichterstattung“ beschrieb er Funktionen der VFU in typischen Workflows der Datennutzung. Das Projekt befindet sich derzeit in der Einführungs- und Entwicklungsphase (2012 – 2014). Anwendungsfall für den Einsatz der VFU ist die Erstellung des Dritten Berichts zur sozio-ökonomischen Entwicklung in Deutschland. Die VFU wird projektbegleitend weiterentwickelt und operativ im Verbund der Projektpartner genutzt. Das Ziel ist die Bereitstellung einer nachhaltigen VFU-Plattform.

Dr. Bartelheimer befasste sich mit der Frage des Transfers der Daten in die VFU, sowie der Abläufe hinter der Datenschnittstelle. Besonderes Augenmerk galt hierbei dem Workflow in der VFU, sowie dem Transfer von Syntax- und Metadateien. Als entscheidend für die Akzeptanz einer VFU beschrieb Dr. Bartelheimer zum einen die Kontrollanforderungen an die datenhaltende Einrichtung, zum anderen die intuitive individuelle Anwendung für den Nutzer. Die VFU sollte eine einheitliche Arbeitsumgebung bieten und es erlauben, schrittweise mehr Daten zu integrieren. Die Forschungsdatenzentren sollten bei der Entwicklung von VFUs und der zugehörigen Nutzungsstudien in die technische und soziale Entwicklung integriert werden.

Reiner Mauer bot in seinem Vortrag einen Blick auf die Persistente Domäne und damit der dauerhaften Bereitstellung von Forschungsdaten. Der Vortrag begann mit einem Überblick über Forschungsdaten in den Sozialwissenschaften und den Aufgaben von GESIS als Infrastruktureinrichtung für die Sozialwissenschaften.

Die Langzeitarchivierung von Forschungsdaten ist ein explizites Ziel von GESIS und als solches in der Satzung verankert. Reiner Mauer positionierte die Aufgaben von GESIS im Kontext des Curation Continuum und beschrieb den Arbeitsablauf bei der Archivierung, der Datenaufbereitung und der Datenbereitstellung. Die Langzeitarchivierung dient der Substanzerhaltung, sowie dem Erhalt von Nutzbarkeit und Interpretierbarkeit. Er betonte, dass das Vertrauen der Nutzer in die dauerhaften Domäne von großer Bedeutung sei – Stichwort Zertifizierung. Herr Mauer erläuterte die Definition und die Aufgaben persistenter Identifikatoren und stellte einige Fragen zu deren Implementierung zur Diskussion.

Als Herausforderungen beschrieb Herr Mauer die immer stärker werdende Bedeutung strukturierender und kooperativer Dienste. Die Bedeutung komplexer Forschungsdesigns und neuer Datenformen wachse, so dass die LZA sich vor der Herausforderung sieht, mit immer komplexeren Forschungsdaten umzugehen. Desweiteren nimmt die Verlinkung bzw. Integration von Daten aus unterschiedlichen Datenquellen und Disziplinen zu. Dies resultiert in höhere Anforderungen an Daten und Metadaten. Rechtliche und Lizenzfragen gewinnen an Bedeutung. Bei der Qualität sind insbesondere Aktualisierung, Versionierung, und autoritative Versionen relevant.

Dr. Hans Pfeiffenberger behandelte in seinem Vortrag Portale und Best Practices und bot damit ein Beispiel für Datennutzung im Kontext der Öffentlichen Domäne. Nach einem blick auf Data Sharing in der Wissenschaft und dem geäußerten Wunsch, über einen singulären Zugangspunkt für möglichst alle wissenschaftlichen Datensätze zu verfügen, ging Dr. Pfeiffenberger näher auf die oftmals sehr unterschiedliche Herkunft der Daten und die Diversität der Datenformen ein. In mehreren Beispielen zeigte er die vielfältigen und verstreut existierenden Datenquellen am Beispiel von Forschungsschiffen und Bojen mit Meßfunktion. Weiterhin stellte er ein Modell zur Klassifizierung und Zuordnung der Daten zu Expeditionen, Forschern und Art der Daten vor. Wichtig sei, alle notwendigen Kontext- und Abstammungsinformationen mit den Daten zu erfassen.

Besonders hervor hob Dr. Pfeiffenberger die Notwendigkeit zur Reproduzierbarkeit der Daten. Er beschrieb den Weg von Rohdaten zu qualitätsgesicherten Primärdaten und daraus abgeleiteten Datenprodukten, und betonte die Notwendigkeit Datenakquise-Praktiken zu harmonisieren.

Workshop-Programm

Die Palette der angebotenen Workshops spiegelte die Bandbreite der von Forschungsdaten-Infrastrukturen angesprochenen Themen wieder. Das Spektrum der Themen reichte von Qualifizierung und Policies über eher anwendungsorientierte Themen wie Software-Lösungen und Datenmanagement bis hin zu rechtlichen Rahmenbedingungen. Die Workshops organisierten sich in drei Themenblöcken „Realisierung“, „Daten“ und „Rahmenbedingungen“. Die Workshops waren organisiert als einstündige Diskussionsrunden mit dem Ziel, offensichtliche Lücken (konzeptionell, organisatorisch, technisch) zu identifizieren, vorbildliche Lösungen aufzuzeigen und eine Vision für das Jahr 2020 (Horizon 2020) zu entwerfen.



Abb. 6: Workshop Impression

Workshop WS 1 – Qualifizierung

Der Workshop wurde eröffnet durch ein Impulsreferat von **Maxi Kindling (Humboldt-Universität zu Berlin)**. In ihrem Impulsreferat erläuterte sie die Integration von Datenmanagement als Thema in fachwissenschaftlichen Studiengängen und gab einen Einblick in die Aus- und Weiterbildungsaktivitäten am Institut für Bibliotheks- und Informationswissenschaft (IBI). Die Workshop-Teilnehmer diskutierten anschließend über die Schaffung eigener Studiengänge für das Datenmanagement, sowie die Integration des Datenmanagements in bibliotheks- und informationswissenschaftliche Studiengänge unter Berücksichtigung des Praxisbezugs. Die Teilnehmer kamen überein, dass die Breite des Qualifizierungsangebots insgesamt erweitert werden müsse. Eine Qualifizierung solle auf den Ebenen des Studiums, des weiterführenden Studiums, sowie als Weiterbildung außerhalb der Universitäten und Fachhochschulen angeboten werden. Der Workshop wurde moderiert von **Prof. Dr. Peter Schirnbacher (Humboldt-Universität zu Berlin)**.

Workshop WS 2A – Datenmanagement

Der Workshop begann mit einem Impulsreferat vom Moderator **Dr. Jens Nieschulze (Universität Göttingen)**. Schwerpunkte der anschließenden Diskussionen waren vor allem Hemmnisse und Anreize für ein gutes Datenmanagement sowie die Nachnutzung der Daten.

Als Hemmnisse wurden unter anderem der erhöhte Zeitaufwand bei der Pflege der Daten für Archivierung und Nachnutzung genannt, die Ungewissheit ob die Daten überhaupt nachgenutzt werden und sich der Aufwand deshalb lohnt, die Form der Daten, welche oftmals nur in analogem Format vorlägen (besonders die Metadaten in Form eines Laborbuches), sowie Geheimhaltungswünsche und -pflichten. Als Anreize wurden die Richtlinien der Forschungsförderer sowie ein „Impaktfaktor“ durch die Publikation von Daten genannt. Allerdings wird ein hoher Nutzen für den Mehraufwand erwartet. Auch erfolgreiche Beispiele werden als ein wichtiger Anreiz gesehen. Lösungen zum Datenmanagement sollten sich nahtlos in die Arbeitsumgebungen der Wissenschaftler integrieren und nachhaltig sein, da sie immer weiterentwickelt werden müssen.

Für die Zukunft wurde der Wunsch nach Datenzentren geäußert, die eng mit bestimmten Communities zusammenarbeiten, miteinander vernetzt sind und eine Weiterbildung im Datenmanagement anbieten. Die Universitäten müssen mehr Routine entwickeln, um mit solchen Datenzentren zusammenzuarbeiten - derzeit werden Datenzentren eher als Zusatzlast empfunden.

Workshop WS 2B – Datenmanagement

Wie kann das Datenmanagement in der privaten bzw. ruppen-Domäne (nach Treloar) besser unterstützt werden? Wie können hier Angebote über punktuelle Lösungen hinaus fruchtbar gemacht und besser koordiniert werden? Welche Aspekte gilt es hier über den der technischen Infrastruktur hinaus zu beachten? Diese Fragen standen im Mittelpunkt des von **Dr. Harry Enke (Leibniz-Institut für Astrophysik, AIP)** moderierten Workshops „Datenmanagement“. Kontrovers diskutiert wurde der Ansatz, Meß- und Beobachtungsdaten bereits bei der Entstehung zu speichern, ohne einen korrigierenden oder erläuternden Eingriff durch den Wissenschaftler abzuwarten. Der Moderator betonte, dass Forschungsdatenmanagement die Begleitung des gesamten Prozesses wissenschaftlichen Arbeitens umfassen müsse und in Datenmanagementplänen festgehalten werden müsse. Auch das Formulieren solcher Pläne könne einen kulturellen Wandel in den Disziplinen beschleunigen, die bislang dem Management der Daten noch nicht die nötige Bedeutung zubilligten.

Workshop WS 3 – Policies

Diskutiert wurde die Vielschichtigkeit von Policies - einerseits als Policies praktischer Natur, welche Handlungsempfehlungen darstellen, andererseits als Policies politischer Natur, welche eine Zielsetzung definieren. Betrachtung in diesem Workshop fanden Policies politischer Natur. In vielen Erklärungen wird davon ausgegangen, dass der Wille zur Wiederverwendung von Forschungsdaten vorhanden ist. Ganz eindeutig ist dies jedoch nicht geklärt, da sich die Gemeinschaft der potentiellen Nachnutzer in zwei Gruppen teilt - den eher observations-orientiert arbeitenden Disziplinen, bei denen die Nachnutzung eine besondere Rolle spielt und den eher experimentell arbeitenden Wissenschaften, bei denen die Beschreibung des Verfahrens eine große Bedeutung hat, um Ergebnisse nachvollziehen zu können. Diskutiert wurde die Frage, ob es legitim sei, eine Veröffentlichung zu fordern und ob Daten in größerem Maß fachübergreifend nachgenutzt werden, da es keine wissenschaftlichen Untersuchungen zu dem Problem gäbe. Eine Nachnutzung von Daten bedeutet immer auch deren Aufbereitung und die Kuratierung der zugehörigen Metadaten - ein teurer Prozess, der zur Differenzierung der Relevanz der Datensätze führt. Der Workshop wurde moderiert von **Dr. Christoph Bruch (Helmholtz Open Access Koordinationsbüro)**.

Workshop WS 4 – Software Lösungen

Der von **Dr. Robert Hauser (FIZ-Karlsruhe)** moderierte Workshop thematisierte Software-Lösungen für Forschungsdateninfrastrukturen (FDI). Bei einer Vorstellung der Teilnehmer zeigte sich, dass viele Institutionen sich gerade in der Evaluationsphase für Software für Forschungsdateninfrastrukturen befinden. In der weiteren Diskussion bestätigte sich die Annahme, dass es keine generelle einzelne Softwarelösung gibt, welche die Anforderungen der unterschiedlichen Fachrichtungen erfüllen kann.

Erkenntnisse des Projektes re3data.org zu existierende Forschungsdatenrepositorien zeigen, dass eine Vielzahl von Lösungen im Einsatz sind, meist Eigenentwicklungen oder Weiterentwicklungen existierender Software. Oft wird der Aufwand von Eigenentwicklungen geringer angesehen als die Integration und Anpassung existierender Lösungen. Wichtige Punkte bei der Entwicklung eigener Software-Lösungen sollten die Zusammenarbeit von Wissenschaftlern und Infrastrukturbetreibern, die Abbildung des gesamten Forschungsprozesses und wissenschaftlichen Workflows in der Software, sowie die Nachhaltigkeit sein. Desweiteren wurde die allgemeine Situation der Entwicklung von Softwarelösungen in Deutschland besprochen. Die Teilnehmer schlugen den Aufbau eines nationalen und zentral gesteuerten Netzwerks von Forschungsdatenzentren vor. Es wurde festgestellt, dass die derzeitigen Akteure und Initiativen (DINI, Wissenschaftsorganisationen, etc.) sich zwar in Netzwerken sammeln, jedoch nicht koordiniert agieren.

Bestehende Initiativen in Deutschland sollten gebündelt werden, um durch gemeinsame Veranstaltungen mit Vorträgen und Experten-Workshops den Austausch zu fördern.

Workshop WS 5A – Datenpublikationen

Der Workshop begann mit einer Präsentation von Moderator **Dr. Michael Lautenschlager (Deutsches Klimarechenzentrum)**, welche den Lebenszyklus der Daten in der Langzeitarchivierung, sowie Daten als Teil virtueller Forschungsumgebungen beschrieb. Laut Moderator gehe der Trend zur Anonymisierung von Datenproduzent und Nutzer. Ziel sei ein interdisziplinäres, verteiltes, selbstbeschreibendes, qualitätsgeprüftes Langzeitdatenarchiv, um Daten ohne Kontakt zum Datenautor nachnutzen zu können. Nachfolgend wurde vor allem über die Thematik der Qualitätssicherung von Daten, sowie der Zertifizierung von Archiven und Prozessen diskutiert.

Workshop WS 5B – Datenpublikationen

Diskutiert wurden die vielen Gesichter der Datenpublikationen, das Thema Qualitätssicherung und Visionen für die Zukunft der Datenpublikationen im Jahr 2020. Gefordert wurden Lektorate für Forschungsdaten, da sich die Bewertung und Qualitätssicherung bislang recht schwierig darstellt. Kritisiert wurden bislang eher unbefriedigende, recht technisch ausgelegte Frameworks für Qualitätskriterien. Desweiteren müsse die Nutzbarkeit und Art der Metadaten bedacht werden. Das Publizieren von Daten wurde als kompliziert bewertet. Moderiert wurde der Workshop von **Roland Bertelmann (Deutsches GeoForschungsZentrum GFZ)**.

Workshop WS 6 – Kosten

Fragestellungen dieses Workshops waren die Kostenermittlung für die Langzeitarchivierung von Forschungsdaten und wie Kosten für das Datenmanagement für einen Projektantrag kalkuliert werden sollten. Desweiteren wurde über potentielle Kostenträger, Preismodelle und noch fehlende Richtlinien zum Datenmanagement für die Förderung von Forschungsaufträgen diskutiert. Moderiert wurde der Workshop von **Dr. Henk Harmsen (DANS)**.

Workshop WS 7 – Persistente Identifier

Der Workshop, moderiert von **Dr. Jens Klump (GFZ)**, wurde durch ein Impulsreferat von **Dr. Janna Neumann (TIB Hannover)** eröffnet. Die Teilnehmer widmeten sich zunächst der Terminologie zur Beschreibung von Identifikatoren. Die Definition des Begriffs Persistente Identifier (PID) ist ausschlaggebend für die Wahl der Anwendungsfelder und die Auswahl des darunter liegenden technischen Dienstes. Die nachfolgende Diskussion gliederte sich in drei Hauptthemen: Die Wahl des Anbieters eines PID-Systems, die Anwendungsfälle von Systemen im Kontext von Forschungsdaten, und die Probleme der Interoperabilität zwischen Systemen und Anbietern.

Eine Klärung der Begriffe war notwendig, um die Frage zu diskutieren, wie noch nicht veröffentlichte Objekte eindeutig referenziert werden, welches System dafür geeignet und ob es überhaupt notwendig ist, für verschiedene Phasen des Lebenszyklus von Forschungsdaten unterschiedliche PID-Systeme zu verwenden. Hinter dieser Frage stehen bisher ungeklärte Anforderungen an Qualität, Vertrauenswürdigkeit und Kontrollierbarkeit der Systeme. Aber auch bei praktischen Fragen der Identität, Granularität und Versionierung digitaler Forschungsobjekte wäre eine Vereinheitlichung der Begriffe hilfreich und würde zur konzeptionellen Klarheit beitragen.

Ein aktuelles Beispiel für unterschiedliche Ansätze für den Betrieb eines PID-Systems ist die Identifikation von Autoren. Hier stehen heute mehrere Systeme zur Auswahl, die teilweise miteinander interoperabel sind. Der Hauptunterschied zwischen den verschiedenen Systemen zur Identifikation von Autoren liegt jedoch in ihrer Organisation und in der Rolle der Verlage und Bibliotheken.

Workshop WS 8A – Daten in Virtuellen Forschungsumgebungen

In diesem Workshop, moderiert von **Dr. Heike Neuroth (Niedersächsischen Staats- und Universitätsbibliothek Göttingen)**, wurden zunächst Definitionen von VFUs besprochen und anschließend Probleme zum Thema VFU aus einer vorgestellten Liste zu priorisiert und Lösungswege besprochen. Im Detail diskutiert wurden der Ablauf des Dialogs zwischen (Fach-)Wissenschaftler und VFU-Betreiber, der Gegensatz von „Einfache Nutzung vs. Alleskönner“, sowie Möglichkeiten zur Sicherstellung der Nachhaltigkeit. Für die Zukunft von VFUs wurde die Wunschvorstellung geäußert, alle Arbeitsabläufe der Forschung so zu integrieren, dass der Wechsel zwischen normaler Arbeitsumgebung und VFU für den Nutzer praktisch nicht merkbar ist. Dafür müsste allerdings die Benutzerfreundlichkeit der VFUs erheblich verbessert werden.

Workshop WS 8B – Daten in Virtuellen Forschungsumgebungen

Zu Beginn des Workshops wurde gemeinsam versucht, Begriffsbestimmungen für Forschungsdaten und Virtuelle Forschungsumgebungen (VFU) zu finden. Die Teilnehmerinnen und Teilnehmer waren sich einig, dass für die Kategorisierung von Forschungsdaten vor allem praktische Gesichtspunkte den Ausschlag geben und dass deren Form, Format und Umfang je nach Wissenschaftsdisziplin sehr unterschiedlich sein können.

Virtuelle Forschungsumgebungen stellen komplexe Systeme dar, die disziplinübergreifend oder disziplinspezifisch verschiedene Tätigkeiten des wissenschaftlichen Arbeitsprozesses abbilden. Dabei sind sie meist kein abgeschlossenes Produkt, sondern ein Framework oder Baukastensystem, das zur Unterstützung der Gesamtheit oder eines Teils des Wissenschaftsprozesses dient. In der weiteren Diskussion wurde die Abbildung der Vorlieben und Bedürfnisse der Nutzer als eine besonders wichtige Anforderung an virtuelle Forschungsumgebungen betont. Bei der institutsübergreifenden Nutzung virtueller Forschungsumgebungen ist es wichtig, wer Zugriff auf die enthaltenen Forschungsdaten haben soll, wer dies festlegt und wie der Zugang organisatorisch und technisch geregelt wird. Wenn Daten öffentlich zugänglich gemacht werden sollen, können Sicherheitsaspekte eine Rolle spielen. Virtuelle Forschungsumgebungen sollten als Dienst flexibel und anpassungsfähig sein um eine möglichst einfache Zusammenarbeit bei der Nutzung der Inhalte zu erreichen. Das Vertrauen der Nutzerinnen und Nutzer ist eine wichtige Voraussetzung für die Annahme virtueller Forschungsumgebungen als Arbeitsmittel und damit auch für die Nachhaltigkeit der entstehenden Strukturen. Bei der Finanzierung müssen gerechte Lösungen für die Umlage der Kosten gefunden werden.

Zusammenfassend wurde festgestellt, dass virtuelle Forschungsumgebungen auch in Zukunft als Abbild des bisherigen Wissenschaftsprozesses ihre Berechtigung haben werden, wenn der Mensch als Nutzer im Mittelpunkt steht. Der Workshop wurde moderiert von **Prof. Dr. Peter Schirnbacher (Humboldt-Universität zu Berlin)**.

Workshop WS 9 – Rechtliche Rahmenbedingungen

Der Umgang mit digitalen Forschungsdaten wird häufig von rechtlichen Fragestellungen berührt. Vor diesem Hintergrund widmete sich der Workshop „Rechtliche Rahmenbedingungen“ der juristischen Dimension der Forschungsdaten-Infrastruktur. Moderiert wurde der Workshop von **John H. Weitzmann (Creative**

Commons). Im Kern der Diskussion standen Fragen wie „Wem gehören (meine) Daten?“ und angrenzende Themen wie die Haftung für Folgeschäden bei „falschen Daten“. Der spannende Workshop zeigte, dass noch viele Fragen ungeklärt sind. Aus Sicht der Wissenschaft scheint der Wunsch nach rechtsicheren Verfahren im Umgang mit Forschungsdaten höchste Priorität zu haben. Zum Abschluss der Diskussion wurde noch ein Blick auf den gesetzgeberischen Handlungsbedarf geworfen. Hier wurde u.a. festgestellt, dass bei den bisherigen Diskussionen über eine weitergehende Wissenschaftsschranke der Themenkomplex Forschungsdaten noch wenig Berücksichtigung fand.



Abb. 7: Die Teilnehmer des FDI 2013-Symposiums

Fazit

In seinem Schlusswort fasste **Dr. Jens Klump (GFZ)** die Ergebnisse des Symposiums zusammen. Mit etwa 140 Teilnehmern, 10 Vorträgen und 12 Workshops war es den einladenden Projekten gelungen, eine große Zahl an Experten zu einem intensiven und anregenden Symposium zusammenzubringen. In den Workshops und am Rande der Veranstaltung konnten die Teilnehmer neue Kontakte knüpfen und so die Vernetzung der Akteure verbessert werden. Die große Zahl der Teilnehmer zeigt, dass das Thema Forschungsdaten als wichtiges Thema wahrgenommen wird. In den Diskussionen war zu beobachten, dass die Entwicklung von Forschungsdaten-Infrastrukturen zwar nach wie vor heterogen verläuft, und auch viele Fragen auch weiterhin unbeantwortet bleiben. Die Diskussion zeigte jedoch auch, dass das Thema Forschungsdaten-Infrastrukturen eine konzeptionelle Reife erreicht hat, die mit Konzepten in anderen europäischen Staaten, den USA oder Australien vergleichbar ist.

Insbesondere in der technischen und der konzeptionellen Entwicklung ist eine gewisse Konvergenz zu beobachten. Daneben zeigen sich in anderen Bereichen konzeptionelle Unschärfen, bei denen noch Forschungsbedarf besteht, so z.B. bei den Themen „Qualität“, „Vertrauen“ und „Kosten- und Preismodelle“. Lücken bestehen auch noch bei Datenmanagement-Werkzeugen und deren Integration in die Arbeitsabläufe der Wissenschaftler. Die Teilnehmer bemängelten, dass viele Akteure im Datenmanagement nicht weit genug bekannt seien und auch die Vernetzung innerhalb der Community optimiert werden könnte. Auch sei der Bedarf an Angeboten in den Bereichen Qualifizierung und Beratung besonders ausgeprägt.

Auch bei dieser Veranstaltung wurde deutlich, dass eine Verbesserung des Umgangs mit Forschungsdaten nicht nur eine Frage der unterstützenden technischen Infrastrukturen ist, sondern auch einen kulturellen

Wandel in der Wissenschaft erfordert. Der kulturelle Wandel im Hinblick auf Forschungsdaten bewegt sich in Richtung eines offeneren Umgangs mit diesem Teil der wissenschaftlichen Überlieferung, wie es im Bericht „*Science as an Open Enterprise*“ der Royal Society (2012) vorgeschlagen wird. Der Grad der Offenheit wird bestimmt durch die Spannung zwischen Vertrauen in die „Peers“ und Kontrolle über das eigene „Werk“. Neben den sozialen Normen muss auch noch der rechtliche Rahmen für Forschungsdaten weiterentwickelt werden.

Für eine Verbesserung der Situation ist es notwendig, die Erstellung von Daten, Software und Infrastrukturen neben den bisher üblichen Literaturveröffentlichungen als Beitrag im wissenschaftlichen Wertesystem zu verankern. Dies bedingt, dass der Nutzen einer Forschungsdateninfrastruktur für Forscher offensichtlich sein muss. Datenpolicies und die Berücksichtigung dieser Leistungen in den institutionellen Bewertungssystemen können diesen Wandel unterstützen.

Positiv bewerteten die Teilnehmer die Möglichkeit, neue und vorbildliche Lösungen für den Umgang mit Forschungsdaten kennenzulernen. Einige Projekte und vielversprechende Ergebnisse wurden vorgestellt. Durch Evaluation der Ergebnisse und Erfahrungsaustausch kann hier eine Entwicklung angestoßen werden, welche über das Stadium des Experimentierens hinausgeht. Wichtig ist, dass der Austausch zwischen den Akteuren weitergeht und auch Praxisvermittlung einschließt, um den Kreis der Akteure zu erweitern.

Sowohl die Organisatoren als auch die Teilnehmer des Symposiums bewerteten die Veranstaltung als äußerst hilfreich für den Austausch von Ideen, der Generierung neuer Impulse und für die Vernetzung der Akteure untereinander. Über Folgeveranstaltungen wird bereits nachgedacht.



Danksagungen

Wir möchten uns herzlich bei unseren Vortragenden, Workshop-Moderatoren und den fast 140 Teilnehmern bedanken, die mit ihren abwechslungsreichen und interessanten Beiträgen unsere Veranstaltung bereichert und neue Impulse für die Weiterentwicklung von Forschungsdaten-Infrastrukturen gegeben haben. Insbesondere danken wir den Protokollanten der Workshops für ihre Mitschriften:

Dr. Torsten Rathmann, Dr. Daniela Koudela, Dr. Beate Rusch, Damian Ulbricht, Paul Vierkant, Vivien Hollad, Dr. Peter Löwe, Britta Dreyer, Markus Schnalke, Dr. Wolfgang Peters-Kottig, Dr. Paul Schulze-Motel und Heinz Pampel.

Lizenz



Alle Texte dieser Veröffentlichung, ausgenommen Zitate, sind unter einem Creative Commons Namensnennung 3.0 Deutschland Lizenzvertrag lizenziert: <http://creativecommons.org/licenses/by/3.0/de>

Zitationsvorschlag

Schäfer, Leonie; Pampel, Heinz; Klump, Jens; Häsler, Tim (2013), Bericht Symposium „Forschungsdaten-Infrastrukturen (FDI 2013)“, Workshop Report, Helmholtz-Zentrum Potsdam Deutsches GeoForschungs-Zentrum, Potsdam, Germany. [online] Available from: http://dx.doi.org/10.2312/RADIESCHEN_003