

re3data.org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositoryn. Ein Werkstattbericht

Heinz Pampel¹, Hans-Jürgen Goebelbecker², Paul Vierkant³

¹ Deutsches GeoForschungsZentrum GFZ, Potsdam

² Karlsruher Institut für Technologie (KIT)

³ Humboldt-Universität zu Berlin

Zusammenfassung

Um Wissenschaftler beim „data sharing“ zu unterstützen, bedarf es u.a. Infrastrukturen mit denen Forschungsdaten zugänglich und nachnutzbar gemacht werden können. Diese Infrastrukturen werden vermehrt unter dem Begriff Forschungsdaten-Repository gefasst. Die Landschaft dieser Repositoryn ist heterogen und unübersichtlich. Die Initiative re3data.org hat das Ziel diese Repositoryn in einem web-basierten Verzeichnis zu erschließen und so eine Orientierung über bestehende Datensammlungen zu bieten. Der Beitrag beschreibt Hintergrund und Anliegen des Projektes.

Abstract

Researchers need support in data sharing, e.g. via infrastructures that allow access and reuse of research data. These infrastructures can be summarised under the term Research Data Repository. The landscape of these repositories is heterogeneous and complex. The aim of re3data.org is to create a web-based registry of Research Data Repositories. The initiative re3data.org will index Research Data Repositories to provide an overview on existing research data collections. The paper describes the background and purpose of the project.

1. Einführung

Die Diskussion um die Zugänglichkeit von Forschungsdaten gewinnt an Relevanz. Die Forderungen internationaler Wissenschaftsorganisationen nach dem „offenen Zugang zu wissenschaftlichem Wissen“ (Berliner Erklärung, 2003) aus dem Jahr 2003 und die von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung 2007 veröffentlichten „Principles and Guidelines for Access to Research Data from Public Funding“ (OECD, 2007) sind zwei von vielen Referenzen einer breiten Debatte, die vielfältige Akteure im System Wissenschaft beschäftigt. Motor dieser Diskussion ist die Digitalisierung von Forschung und Lehre. Die Forderung nach „langfristige[r] Sicherung und [...] grundsätzlich[em] offenen Zugang zu Daten aus öffentlich geförderter Forschung“, die die deutschen Wissenschaftsorganisationen 2010 in „Grundsätze zum Umgang mit Forschungsdaten (Allianz 2010) formuliert haben, kann als Mission der digitalen Wissenschaft

verstanden werden. Ziel ist es, die „Produktivität und Wettbewerbsfähigkeit der Wissenschaft“ (Allianz 2010) zu fördern, indem die Nachprüfbarkeit (z.B. im Rahmen der Qualitätssicherung) und die Nachnutzung (z.B. zur Vermeidung von Doppelerhebungen der Daten) ermöglicht werden.

Eine anhaltende Herausforderung stellt eine interdisziplinär geltende Definition des Begriffes Forschungsdaten dar. Im Folgenden werden unter diesem Begriff digitale Daten gefasst, die je nach Disziplin in unterschiedlichsten Formen und Formaten vorliegen. Diese können während eines Forschungsprozesses entstanden sein oder Grundlage dieses Prozesses sein.¹

Verstärkt wirken sogenannte „Data Policies“ auf Wissenschaftler und deren Umgang mit Forschungsdaten.² Hervorzuheben sind die empfehlenden und verpflichtenden Anforderungen von Förderorganisationen und wissenschaftlichen Zeitschriften, die Mittelempfänger und Autoren auffordern, die Zugänglichkeit der Daten, die im Rahmen eines Projektes entstehen oder Grundlage einer Publikation sind, sicher zu stellen. So sind z. B. Antragsteller der National Science Foundation (NSF) durch eine „Data Sharing Policy“ gefordert, „to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.“ (NSF, 2011a) Weiter fordert die NSF, dass Maßnahmen zur Umsetzung dieser Richtlinie in einem „Data Management Plan“ spezifiziert werden (NSF, 2011b). Auch die Deutsche Forschungsgemeinschaft (DFG) erwartet seit 2010 im Rahmen ihres Programms „Sachbeihilfe“ Aussagen zu Verfahren, „um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen“. Dabei sollen „sofern vorhanden [...] Angebote bestehender Datenrepositorien“ berücksichtigt werden (DFG, 2012). Dieses Anliegen wird vermehrt auch in den „Editorial Policies“ von wissenschaftlichen Zeitschriften deutlich. Beispielhaft seien hier die Anforderungen der Nature-Zeitschriftenfamilie genannt. In diesen heißt es: „authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications in material transfer agreements“. Angeregt wird die Zugänglichmachung der Daten „via public repositories.“ (Nature, 2011).

¹ Diese Definition wurde abgeleitet von Kindling (2009) und Kommission Zukunft der Informationsinfrastruktur (2011).

² Siehe dazu ausführlich: Pampel & Bertelmann (2011).

Vor dem Hintergrund der skizzierten Entwicklungen manifestieren sich Strategien zur Zugänglichmachung von Forschungsdaten. Kategorisierungen dieser Publikationsstrategien finden sich bei Dallmeier-Tiessen (2011) und Lawrence et al. (2011). Angelehnt an Dallmeier-Tiessen können drei Strategien der Zugänglichmachung unterschieden werden: 1. Die Veröffentlichung von Forschungsdaten als eigenständiges Informationsobjekt in einem Repositorium. 2. Die Veröffentlichung von Forschungsdaten mit textueller Dokumentation, als sogenanntes „Data Paper“³. 3. Die Veröffentlichung von Forschungsdaten als Anreicherung einer interpretativen Text-Publikation („Enriched Publication“⁴). Gemein ist diesen Publikationsstrategien, dass es einer Infrastruktur bedarf, die die Speicherung und Zugänglichmachung der Daten möglichst dauerhaft sicherstellt. Diese Infrastrukturen werden im Folgenden unter dem Begriff Forschungsdaten-Repositorien gefasst.

2. Forschungsdaten-Repositorien

Die Europäische Kommission stellte 2009 fest: „The landscape of data repositories across Europe is fairly heterogeneous, but there is a solid basis to develop a coherent strategy to overcome the fragmentation and enable research communities to better manage, use, share and preserve data.“ (European Commission, 2009). Das Zitat beschreibt die heutige unübersichtliche Landschaft von Infrastrukturen, die Forschungsdaten speichern und zugänglich machen. Forschungsdaten-Repositorien sind durch die Daten geprägt, die sie speichern. Anders als Text-Repositorien, die mehrheitlich Informationsobjekte, die im PDF-Format gespeichert sind, zugänglich machen und sich im Kontext von Open Access etabliert haben, speichern Forschungsdaten-Repositorien vielfältige Dateiformate.⁵

Bisher haben sich nur wenige Untersuchungen der weltweiten Landschaft von Forschungsdaten-Repositorien angenommen: Marcial & Hemminger beschreiben in einer 2010 veröffentlichten Studie die Charakteristik von 100 Repositorien, einen

³ Siehe hierzu Chavan & Penev (2011).

⁴ Eine Beschreibung des Konzeptes der „Enriched Publication“ findet sich bei Woutersen-Windhouwer & Brandsma (2009).

⁵ Während sich Anfang der 2000er Jahre die Open Archives Initiative (OAI) formierte, die die Standardisierung und Vernetzung von Repositorien, die textuelle Informationsobjekte (wie Qualifikationsarbeiten und Pre- und Post-Prints) im Sinne des Open Access zugänglich machen, förderte, steht eine vergleichbare Entwicklung im Bereich der Forschungsdaten-Repositorien noch aus.

ähnlichen Ansatz verfolgt Schaaf 2011. Erhellend ist der Blick auf disziplinäre Untersuchungen. Eine eindrucksvolle Landschaft an Daten-Repositoryn bietet die Biomedizin. Die 2012-Ausgabe der „Molecular Biology Database Collection“ des biomedizinischen Journals „Nucleic Acids Research“ weist 1.380 Infrastrukturen nach, auf denen biomedizinische Forschungsdaten gespeichert werden können (Galperin & Fernández-Suárez, 2012). Im Rahmen des Projektes European Life Science Infrastructure for Biological Information (ELIXIR) wurden 200 dieser Infrastrukturen näher untersucht. Diese 200 Repositoryn werden von 100 Institutionen betrieben. Insgesamt sind 350 Personen in den Betrieb der Repositoryn involviert, die durch mehrere hunderttausend Wissenschaftler pro Jahr etwa 60 Millionen Zugriffe verzeichnen. Die jährlichen Betriebskosten für diese 200 Repositoryn liegen bei etwa 30 Millionen Euro (ELIXIR, n.d.). Um den nachhaltigen Betrieb dieser Repositoryn-Landschaft zu garantieren, wurde ELIXIR in das „European Strategy Forum on Research Infrastructures (ESFIR)“ aufgenommen. Anliegen von ESFIR ist die strategische Förderung von Forschungsinfrastrukturen, die für die Wettbewerbsfähigkeit des Europäischen Forschungsraums (ERA) von zentraler Bedeutung sind. Bereits im ESFRI-Einsatzantrag des Rats der Europäischen Union wird deutlich, dass Forschungsinfrastrukturen heute keinesfalls nur physische Infrastrukturen wie Forschungsschiffe oder Beschleuniger sind, sondern auch digitale Infrastrukturen, wie z.B. „electronic archiving systems for scientific publications and databases“ (ESFRI, 2006).

Eine grobe Systematisierung dieser Repositoryn in institutionelle, disziplinäre, multidisziplinäre und projektspezifische Infrastrukturen ist möglich. Beispiel für ein institutionelles Forschungsdaten-Repository ist Edinburgh DataShare.⁶ Dieses „online digital repository of multi-disciplinary research datasets produced at the University of Edinburgh“ (Edinburgh DataShare, o.J.) wird mit der Software DSpace realisiert und wurde im Zeitraum 2007-2009 aufgebaut (Rice, 2009). Im Juni 2012 sind 19 Datensätze auf dem Repository gespeichert.

Ein weiteres Beispiel für ein solches institutionelles Daten-Repository ist Open Data LMU⁷, dieses 2010 gestartete Repository „steht [...] allen Wissen-

⁶ <http://datashare.is.ed.ac.uk>

⁷ <http://data.ub.uni-muenchen.de>

schaftler/innen der LMU⁸ als Publikationsplattform für Primärdaten“ zur Verfügung (Schallehn, 2010). Im Juni 2012 sind 34 Datensätze auf dem Repository gespeichert. Verwendet wird die Software EPrints.

Im Bereich der disziplinären Forschungsdaten-Repositoryn sind Beispiele wie GenBank oder PANGAEA - Data Publisher for Earth & Environmental Science⁹ herauszuheben. GenBank¹⁰, wurde 1982 ins Leben gerufen¹¹ und definiert sich als „public database of nucleotide sequences and supporting bibliographic and biological annotation.“ Betrieben wird diese Infrastruktur, die Nukleotidsequenzen von mehr als 250.000 Spezies nachweist, vom National Center for Biotechnology Information (NCBI) (Benson et al., 2012). PANGAEA versteht sich als „Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research.“ (PANGAEA, n.d.). Dieses Repository wird vom Alfred-Wegener-Institut für Polar- und Meeresforschung (AWI) und vom Zentrum für Marine Umweltwissenschaften der Universität Bremen (MARUM) betrieben. Aufgebaut wurde PANGAEA im Rahmen des Projektes „Paläoklima-Datenzentrum“, das vom Bundesministerium für Bildung und Forschung (BMBF) von 1994 bis 1997 gefördert wurde (Diepenbroek et al., 1999). 2011 speichert PANGAEA über eine halbe Million geowissenschaftlicher Datensätze (Schindler, Diepenbroek, & Grobe, 2012).

Einige Infrastrukturen verfolgen auch einen multidisziplinären Ansatz, z.B. FigShare¹², eine 2011 gestartete „permanent research data storage and sharing platform“ der Verlagsgruppe Macmillan, oder LabArchives¹³, eine mit dem Verlag BioMed Central kooperierende privatwirtschaftlich organisierte „platform for publishing scientific data“ (Hrynaszkiewicz, 2012).

Vielfältig ist auch die Landschaft der projektspezifischen Repositoryn. Exemplarisch kann hier die vom Deutsches GeoForschungsZentrum GFZ betriebene Scientific Drilling Database (SDDDB)¹⁴ genannt werden, die Bohrdaten, die im Rahmen des Scientific Continental Drilling Program (ICDP) entstehen, offen zugänglich macht (Klump & Conze, 2007). Ein weiteres Beispiel ist das Repository des Bern Digital

⁸ Ludwig-Maximilians-Universität München

⁹ <http://www.pangaea.de>

¹⁰ <http://www.ncbi.nlm.nih.gov/genbank>

¹¹ Zur Geschichte von GenBank siehe: Cravedi (2008).

¹² <http://figshare.com>

¹³ <http://www.labarchives.com>

¹⁴ <http://www.scientificdrilling.org>

Pantheon Project¹⁵, in dem hochauflösende Fotografien und maßstabsgetreue Visualisierungen des Pantheon in Rom frei zugänglich gemacht wurden. Darüber hinaus gibt es viele weitere Typen von Forschungsdaten-Repositoryn, die teils nur schwer zu charakterisieren sind.

Betrachtet man die zukünftige Bedeutung dieser Infrastrukturen für eine digital arbeitende Wissenschaft, muss eine mangelnde Forschung und Entwicklung konstatiert werden. Der wissenschaftspolitische Wille einer offenen Wissenschaft (Kroes, 2012a; Kroes, 2012b) ist ohne verlässliche sowie vertrauenswürdige und damit auch dauerhafte Infrastrukturen, die Wissenschaftler beim „sharing“ ihrer Daten unterstützen, zum Scheitern verurteilt. Befragungen von Betreibern von Forschungsdaten-Repositoryn zeigen, dass insbesondere die finanzielle Absicherung dieser Infrastrukturen für längere Zeiträume (über 5 Jahre) unsicher ist (Kuipers & Van der Hoeven, 2009).

3. re3data-org: Ein Werkstattbericht

Im Auftrag der Europäischen Kommission wurde 2010 eine Vision des Umgangs mit Forschungsdaten im Jahr 2030 beschrieben. Nach dieser sollen Wissenschaftler in Zukunft in der Lage sein, „to find, access and process the data they need.“ Weiter sollen Forschende, die Daten erheben, diese zukünftig in verlässlichen und vertrauenswürdigen Repositoryn speichern, die auf Basis internationaler Standards arbeiten (High Level Expert Group on Scientific Data, 2010). Mit Blick auf die heterogene Landschaft der Forschungsdaten-Repositoryn ist die Umsetzung dieser Vision eine zentrale Herausforderung für die digital arbeitende Wissenschaft.

Vor dem Hintergrund der skizzierten Forderung nach „data sharing“, die in der jüngsten Vergangenheit vermehrt auf die wissenschaftliche Praxis wirkt (siehe Abschnitt 1), und der heterogenen Landschaft der Forschungsdaten-Repositoryn (siehe Abschnitt 2) verfolgt das Projekt re3data.org - Registry of Research Data Repositories¹⁶ den Aufbau eines Verzeichnisses von Forschungsdaten-Repositoryn.

¹⁵ <http://www.digitalpantheon.ch/repository>

¹⁶ <http://re3data.org>

Die Initiative hat die Erfassung und den Nachweis von Forschungsdaten-Repositoryn aller Disziplinen in einem zentralen web-basierten Recherche- und Nachweissystem zum Ziel.

Das Projekt wird im Zeitraum 2012-2014 von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Projektpartner sind die Abteilung Bibliothek und Informationsdienste des Deutschen GeoForschungsZentrums GFZ¹⁷, das Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin¹⁸, sowie die Bibliothek des Karlsruher Instituts für Technologie (KIT)¹⁹. Vernetzungspartner ist die Deutsche Initiative für Netzwerkinformation (DINI)²⁰, in deren Kontext die drei Projektpartner seit längerem rund um den Betrieb von Repositoryn zusammenarbeiten und 2009 ein Positionspapier zum Thema Forschungsdaten veröffentlicht haben (Dallmeier-Tiessen et al., 2009).

Zentrales Anliegen des Projektes 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 Zeitschriften aufgefordert sind Daten zugänglich machen, als auch die Rolle der Nutzenden, z.B. Forschende, die Daten von Dritten nachnutzen möchten, fokussiert. Weiter soll Förderorganisationen und infrastrukturellen Dienstleistern, wie Datenzentren, Rechenzentren und Bibliotheken, eine Übersicht über die Landschaft der Forschungsdaten-Repositoryn gegeben werden. Darüber hinaus wird das Ziel verfolgt, einen Beitrag zur Schaffung eines „ecosystem of data repositories“ (Van der Graaf & Waaijers, 2011) zu leisten. Das Verzeichnis soll auch den Entwicklungsstand der Repositoryn abbilden und kann so z. B. auch bei der Identifikation von Disziplinen helfen, in denen es eine lückenhafte Repositoryn-Landschaft gibt. Übergeordnetes Ziel von re3data.org ist es, einen Beitrag zur Förderung des „data sharings“ in der Wissenschaft zu leisten.

In der ersten Phase des Projektes steht, ausgehend von der konzeptionellen Entwicklung des Verzeichnisses, die Recherche und Erschließung bestehender Forschungsdaten-Repositoryn, sowie die technische Realisierung des Verzeichnisses im Fokus.

¹⁷ <http://www.gfz-potsdam.de>

¹⁸ <http://www.ibi.hu-berlin.de>

¹⁹ <http://www.bibliothek.kit.edu>

²⁰ <http://www.dini.de>

Während bestehende Listen von Forschungsdaten-Repositoryn mehrheitlich nur die Disziplin und den Betreiber eines Repositorys auszeichnen²¹, ist re3data.org bestrebt, eine möglichst umfassende Beschreibung eines Repositorys zu bieten und dabei z.B. die unterschiedlichen Zugangsbedingungen zu dem Dienst und den darin vorgehaltenen Forschungsdaten darzustellen. Dieses Anliegen einer umfassenden Darstellung von Forschungsdaten-Repositoryn prägt die Konzeption des Verzeichnisses und wirkt auf dessen technische Realisierung.

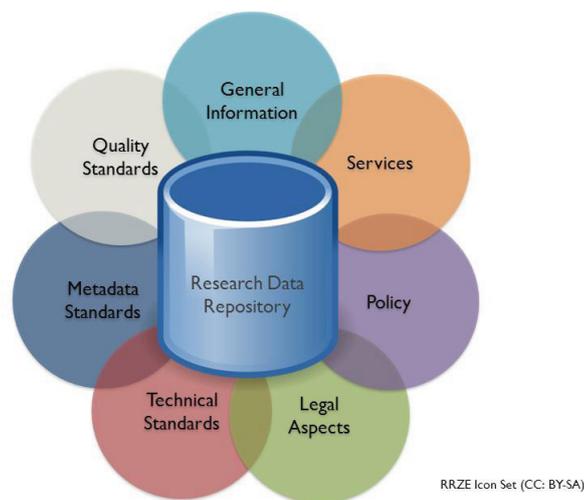


Abbildung 1: Aspekte eines Repositorys, die im Vokabular (Version 1.0) von re3data.org betrachtet werden.

Ausgehend von einer Erhebung von rund 400 Daten-Repositoryn wurde ein Vokabular zur Beschreibung von Forschungsdaten-Repositoryn auf Basis einer Untersuchung, nach zuvor entwickelten Kriterien, von 20 zufällig gewählten Repositoryn entwickelt (Vierkant et al., 2012).

Bei der Erhebung wurde auf bestehende Sammlungen zurückgegriffen, die durch eigene Recherchen ergänzt wurden. Weiter wurde das Projekt, nach der Kommunikation des Projektvorhabens im April 2012, von verschiedenen Repositoryn-Betreibern über ggf. zu erschließende Dienste informiert.

²¹ Siehe z.B. die auf der Website von DataCite integrierte Liste von Repositoryn (<http://datacite.org/repolist>), oder die im Open Access Directory dokumentierten Repositoryn (http://oad.simmons.edu/oadwiki/Data_repositories).

Das Vokabular gliedert sich in seiner Version 1.0 in folgende Bereiche: Allgemeine Informationen, Verantwortlichkeiten, Policies, rechtliche Aspekte, technische Standards, Metadaten-Standards, Qualitätsstandards und Services (siehe Abb. 1). Dank des Vokabular ist es zum Beispiel möglich in re3data.org abzufragen, welches zertifizierte Forschungsdaten-Repositorium einem Datenproduzenten die Speicherung von geowissenschaftlichen Daten unter einer Creative-Commons-Lizenz erlaubt.

Im Kontext der Qualitätsstandards wurden bestehende Anforderungen an Repositorien im Kontext der Zertifizierung eruiert (u.a. Braun et al., 2010; CCSSS, 2011; Data Seal of Approval, 2010; ESF & EUROHORCs, 2011; ICSU WDS, 2011).²² Um der heterogenen Landschaft und den vielfach fehlenden Standards im Bereich der Forschungsdaten-Repositorien zu entsprechen, sind die formulierten Anforderungen an Repositorien, die in dem Verzeichnis indexiert werden, niedrigschwellig gehalten. Neben Pflichtangaben bei der Erschließung der Forschungsdaten-Repositorien sind Angaben zu Zugang²³ und Lizenzierung der Forschungsdaten Mindestanforderungen, die Repositorien-Betreiber erfüllen müssen, um in re3data.org aufgenommen zu werden.

Zur technischen Realisierung des Verzeichnisses werden verschiedene Softwaresysteme zum Betrieb von digitalen Bibliotheken eruiert. Eine bereits in der Antragsphase von re3data.org realisierte Testversion von re3data.org wird erweitert.

Die Entwicklung von re3data.org soll im Dialog mit etablierten Akteuren im Bereich des Forschungsdatenmanagements betrieben werden. Vor diesem Hintergrund wurde im Frühjahr 2012 ein Memorandum of Understanding mit DataCite²⁴ unterzeichnet. DataCite ist eine aus dem von der DFG-Projekt "Publikation und Zitierbarkeit wissenschaftlicher Primärdaten" (STD-DOI)²⁵ entstandene Initiative zur dauerhaften Adressierung von Forschungsdaten. Im Rahmen der Zusammenarbeit soll insbesondere der Informationsfluss zwischen den beiden Partnern gesichert werden.

²² Siehe hierzu auch die Einführung von Klump (2011) sowie die im Rahmen des EU-Projektes APARSEN entstandenen Reports rund um die Qualitätssicherung von Forschungsdaten, z.B. Pampel et al. (2012).

²³ Zugang meint hier den Zugang zum Forschungsdaten-Repositorium, den darin vorgehaltenen Forschungsdaten sowie den Zugang zum Daten-Upload.

²⁴ <http://datacite.org>

²⁵ Eine Fortsetzung der Kooperation der damals beteiligten Partner wird seit August 2011 im Rahmen des DFG-Projektes „Kompetenzzentrum für Forschungsdaten aus Erde und Umwelt“ (KomFor) umgesetzt. Siehe dazu: <http://www.komfor.net>.

4. Fazit

„We start the era of open science.“ Dieses Zitat der EU-Kommissarin für die Digitale Agenda Neelie Kroes (2012a) macht deutlich, dass Offenheit zum Paradigma der digitalen Wissenschaft wird. Zur Förderung dieser Entwicklung wird eine dauerhafte Informationsinfrastruktur benötigt, die Wissenschaftler beim „sharing“ ihrer Daten unterstützt und auch kommenden Forschergenerationen Zugang und Nachnutzung der Daten garantiert. Mitte 2012 befindet sich re3data.org nach einer sechsmonatigen Konzeptionsphase in der Realisierungsphase. Im Fokus steht dabei die technische Umsetzung des Verzeichnisses. Über den Aufbau des Verzeichnisses hinaus, widmet sich das Projekt der Vernetzung und Standardisierung von Forschungsdaten-Repositoryn und leistet so einen Beitrag auf dem herausfordernden Weg zu einer offenen Wissenschaft.

Literaturverzeichnis

Alle Links wurden am 09.07.2012 geprüft.

- Allianz der deutschen Wissenschaftsorganisationen. (2010, June). Grundsätze zum Umgang mit Forschungsdaten. Retrieved from <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsätze/>
- Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 40(Database issue), D48-D53. doi:10.1093/nar/gkr1202
- Berliner Erklärung. (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Retrieved from <http://oa.mpg.de/lang/de/berlin-prozess/berliner-erklarung/>
- Braun, K., Buddenbohm, S., Dobratz, S., Herb, U., Müller, U., Pampel, H., Schmidt, B., et al. (2010). *DINI-Zertifikat Dokumenten- und Publikationsservice 2010* (3.0 ed.). Göttingen. doi:urn:nbn:de:kobv:11-100109986
- Chavan, V., & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, 12(Suppl 15), S2. doi:10.1186/1471-2105-12-S15-S2
- Consultative Committee for Space Data Systems (CCSDS). (2011). *Audit and Certification of Trustworthy Digital Repositories. Recommended Practice. CCSDS 652.0-M-1. Magenta Book. Practice*. Retrieved from <http://public.ccsds.org/publications/archive/652x0m1.pdf>
- Cravedi, K. (2008). *GenBank Celebrates 25 Years of Service with Two-Day Conference. Leading Scientists Will Discuss the DNA Database at April 7-8 Meeting*. Retrieved from <http://www.nih.gov/news/health/apr2008/nlm-03.htm>

- Dallmeier-Tiessen, S, Dobratz, S., Gradmann, S., Horstmann, W., Kleiner, E., Pampel, H., Schirmbacher, P., et al. (2009). *Positionspapier Forschungsdaten* (1.0 ed.). doi:urn:nbn:de:kobv:11-10098082
- Dallmeier-Tiessen, Sünje. (2011). *Strategien bei der Veröffentlichung von Forschungsdaten*. Berlin. Retrieved from http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_173.pdf
- Data Seal of Approval. (2010). *Data Seal of Approval. Guidelines version 1*. Retrieved from <http://assessment.datasealofapproval.org/documentation/>
- Deutsche Forschungsgemeinschaft. (2012). *Leitfaden für die Antragstellung. DFG-Vordruck 54.01 - 1/12*. Retrieved from http://www.dfg.de/formulare/54_01/54_01_de.pdf
- Diepenbroek, M., Grobe, H., Reinke, M., Schlitzer, R., & Sieger, R. (1999). Data management of proxy parameters with PANGAEA. In G. Fischer & G. Wefer (Eds.), *Use of Proxies in Paleoceanography. Examples from the South Atlantic* (pp. 715–727). Berlin, Heidelberg & New York: Springer-Verlag. Retrieved from <http://hdl.handle.net/10013/epic.11224>
- ELIXIR. (n.d.). *The ELIXIR Strategy for Data Resources. Draft Report from Workpackage 2. The ELIXIR Preparatory Phase*. Retrieved from <http://www.elixir-europe.org/bcms/elixir/Documents/reports/ELIXIR Strategy for Data Resources report.pdf>
- ESF & EUROHORCS. (2011). *Basic Requirements for Research Infrastructures in Europe*. Retrieved from http://www.dfg.de/download/pdf/foerderung/programme/wgi/basic_requirements_research_infrastructures.pdf
- Edinburgh DataShare. (n.d.). *What is Edinburgh DataShare?* (pp. 6-8). Retrieved from <http://datashare.is.ed.ac.uk>
- European Commission. (2009). ICT infrastructures for e-science. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM(2009) 108 final. Retrieved from <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF>
- European Strategy Forum on Research Infrastructures. (2006). *European Roadmap for Research Infrastructures. Report 2006*. Retrieved from ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_en.pdf
- Galperin, M. Y., & Fernández-Suárez, X. M. (2012). The 2012 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Research*, 40(Database issue), D1-D8. doi:10.1093/nar/gkr1196
- High Level Expert Group on Scientific Data. (2010). *Riding the wave. How Europe can gain from the rising tide of scientific data*. Retrieved from <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>
- Hrynaszkiewicz, I. (2012, April). LabArchives and BioMed Central: a new platform for publishing scientific data. *BioMed Central Blog*. Retrieved from http://blogs.openaccesscentral.com/blogs/bmcblog/entry/labarchives_and_biomed_central_a
- ICSU World Data System (WDS). (2011). *Certification of World Data System Members Introduction*.
- Kindling, M. (2009). *Möglichkeiten der Strukturmodellierung: Eine exemplarische Zusammenführung funktionaler Anforderungen an die Bereitstellung digitaler Forschungsdaten für ausgewählte geisteswissenschaftliche Disziplinen*. Humboldt-Universität zu Berlin. Retrieved from <http://www.nbn-resolving.org/urn:nbn:de:kobv:11-100185124>

- Klump, J. (2011). Criteria for the Trustworthiness of Data Centres. *D-Lib Magazine*, 17(1/2). doi:10.1045/january2011-klump
- Klump, J., & Conze, R. (2007). The Scientific Drilling Database (SDDB) - Data from Deep Earth Monitoring and Sounding. *Scientific Drilling*, (4), 30-31. doi:10.2204/iodp.sd.4.06.2007
- Kommission Zukunft der Informationsinfrastruktur. (2011). *Gesamtkonzept für die Informationsinfrastruktur in Deutschland*. Retrieved from http://www.allianz-initiative.de/fileadmin/user_upload/KII_Gesamtkonzept.pdf
- Kroes, N. (2012a). *Opening Science Through e-Infrastructures*. SPEECH/12/258. Retrieved from europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258
- Kroes, N. (2012b). *Making Open Access a reality for Science*. SPEECH/12/392. Retrieved from <http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/392&format=HTML&aged=0&language=EN&guiLanguage=en>
- Kuipers, T., & Van der Hoeven, J. (2009). *Insight into digital preservation of research output in Europe. Survey Report*. Retrieved from http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Lawrence, B., Jones, C., Matthews, B., Pepler, S., & Callaghan, S. (2011). Citation and Peer Review of Data: Moving Towards Formal Data Publication. *International Journal of Digital Curation*, 6(2). doi:10.2218/ijdc.v6i2.205
- Marcial, L. H., & Hemminger, B. M. (2010). Scientific Data Repositories on the Web : An Initial Survey. *Journal of the American Society for Information Science*, 61(10), 2029-2048. doi:10.1002/asi
- National Science Foundation. (2011a). *Award and Administration Guide. Chapter VI - Other Post Award Requirements and Considerations*. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4
- National Science Foundation. (2011b). *Grant Proposal Guide. Chapter II - Proposal Preparation Instructions*. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp
- Nature. (2011). *Guide to Publication Policies of the Nature Journals*. Retrieved from <http://www.nature.com/authors/gta.pdf>
- Organisation for Economic Co-operation and Development (OECD). (2007). *Principles and Guidelines for Access to Research Data from Public Funding*. Paris. Retrieved from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- PANGAEA. (n.d.). About / Imprint. Retrieved July 29, 2011, from <http://www.pangaea.de/about/>
- Pampel, H., & Bertelmann, R. (2011). „Data Policies“ im Spannungsfeld zwischen Empfehlung und Verpflichtung. In S. Büttner, H.-C. Hobohm, & L. Müller (Eds.), *Handbuch Forschungsdatenmanagement* (pp. 49-61). Bad Honnef: Bock + Herchen. doi:urn:nbn:de:kobv:525-opus-2287
- Pampel, H., Pfeiffenberger, H., Schäfer, A., Smit, E., Pröll, S., & Bruch, C. (2012). *Report on Peer Review of Research Data in Scholarly Communication*. Retrieved from <http://epic.awi.de/30353/>
- Rice, R. (2009). *DISC-UK DataShare*. Retrieved from <http://ie-repository.jisc.ac.uk/336/>
- Schaaf, I. (2011). *Forschungsdaten im Netz – Untersuchung von online verfügbaren Repositorien*. Hochschule der Medien Stuttgart.

- Schallehn, V. (2010, October 13). Open Data LMU. *Universitätsbibliothek München*. Retrieved from http://www.ub.uni-muenchen.de/no_cache/aktuelles/einzelne-nachricht/article/open-data-lmu/
- Schindler, U., Diepenbroek, M., & Grobe, H. (2012). PANGAEA - Research Data enters Scholarly Communication. *Geophysical Research Abstracts*, 14, EGU2012-13378-1. Retrieved from <http://meetingorganizer.copernicus.org/EGU2012/EGU2012-13378-1.pdf>
- Vierkant, P., Spier, S., Rücknagel, J., Gundlach, J., Kindling, M., Pampel, H., Manova, S., et al. (2012). *Vocabulary for the Registration and Description of Research Data Repositories. Version 1.0*. doi:10.2312/re3.001
- Woutersen-Windhouwer, S., & Brandsma, R. (2009). Enhanced Publications, State of the Art. In M. Vernooy-Gerritsen (Ed.), *Enhanced Publications. Linking Publications and Research Data in Digital Repositories*. Amsterdam: Amsterdam University Press.
- Van der Graaf, M., & Waaijers, L. (2011). *A Surfboard for Riding the Wave. Towards a four country action programme on research data*. Retrieved from http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=/Files/Filer/downloads/Primary+Research+Data/Surfboard+for+Riding+the+Wave/KE_Surfboard_Riding_the_Wave_Screen.pdf