



Originally published as:

Klump, J., Bertelmann, R. (2013): Forschungsdaten - In: Kuhlen, R., Semar, W., Strauch, D. (Eds.), Grundlagen der praktischen Information und Dokumentation, De Gruyter, 575-583

Jens Klump, Roland Bertelmann

D 8 Forschungsdaten

In den vergangenen Jahren war immer wieder die Rede von einer Flut von Daten, deren Bewältigung die Wissenschaft vor immer neue Herausforderungen stellen werde (Lit. 01, Lit. 02). In der Tat hat die schnell wachsende Verfügbarkeit von Daten, Speicher- und Rechenkapazität dazu geführt, dass sich heute neben Empirie, Theorie und Simulation nun auch die Untersuchung großer Datenbestände zu einem eigenen Forschungsansatz entwickelt (Lit. 03). *Big Data* hat sich dabei schnell als treffendes Bild für umfangreiche Volumina etabliert. In vielen Fällen sind die Datenvolumen allerdings vergleichsweise gering, viele Datenbanken sind nur einige Gigabyte groß und wurden daher oft *small data* genannt (Lit. 04, Abbildung 1). Es ist jedoch nicht die schiere Masse der Daten, sondern die Tatsache, dass Daten die Grundlage für neue Erkenntnisse bieten, weshalb dieser neue Ansatz auch *data intensive science* (Lit. 05) oder *data driven science* genannt wird.

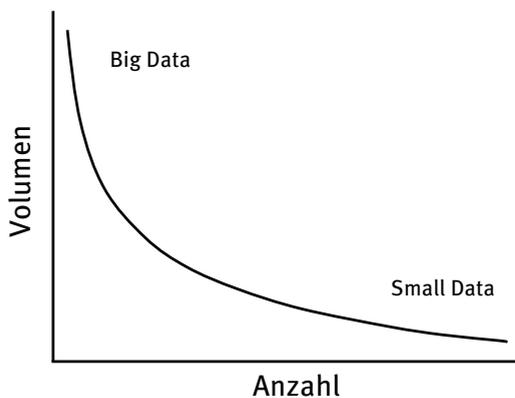


Abb. 1: Die Anzahl der Datensätze des „Big Data“ ist klein, die meisten Forschungsdaten sind „Small Data“.

In einer Reihe von Veröffentlichungen und Positionspapieren haben Forscher (Lit. 06, Lit. 07, Lit. 08), Fachgesellschaften (Lit. 09), Fachzeitschriften und Wissenschaftsorganisationen (Lit. 10, Lit. 11, Lit. 12) auf die Bedeutung von Forschungsdaten für das Erlangen neuer wissenschaftlicher Erkenntnisse hingewiesen. Zwar stellen sie die empirische Grundlage des wissenschaftlichen Prozesses dar, dennoch führten sie in den vergangenen zwei Jahrzehnten ein Schattendasein in der wissenschaftlichen Kommunikation und -überlieferung. War es bis in die achtziger Jahre des 20. Jahrhunderts noch üblich Daten als Tabellen in wissenschaftlichen Veröffentlichungen abzudrucken, so führte die enorme Zunahme der verwendeten Daten und der veröffentlichten Fachartikel sowie eine veränderte Publikationskultur dazu, dass kaum noch Daten zusammen mit dem sie interpretierenden Text veröffentlicht wurden (Lit. 13, Lit. 14). Zudem besteht im Zeitalter digitaler Medien nicht mehr die Notwendigkeit, papierbasierte Medien digital zu emulieren (Lit. 15).

Ausgehend von der Prämisse, dass Forschungsdaten nachgenutzt werden sollen, wollen wir in diesem Kapitel den Lebenszyklus der Daten vom Zeitpunkt der erneuten Nutzung bis zu ihrer Entstehung zurückverfolgen. Unser Modell des Datenlebenszyklus orientiert sich am Modell, das von DataONE veröffentlicht wurde (Lit. 16). Abbildung 2 illustriert aus der Sicht des Forschers wie verschiedene Phasen des Lebenszyklus von Forschungsdaten aufeinander folgen. Die hier benannten Phasen sind die Kernbereiche des Umgangs mit Forschungsdaten und sind Grundlage der Gliederung dieses Kapitels. Nachfolgend werden wir den Umgang mit Forschungsdaten von der Suche nach Daten (*discovery*) aus entlang dieses Referenzmodells zurückverfolgen. Es sei an dieser Stelle auch angemerkt, dass der Umgang mit Forschungsdaten sich inzwischen zu einem eigenen Arbeits-

feld entwickelt hat, so dass dieses Kapitel nur einen Überblick bieten kann. Wir verweisen daher auf die Literatur für weiterführende Information (z. B. Lit. 17, Lit. 18).

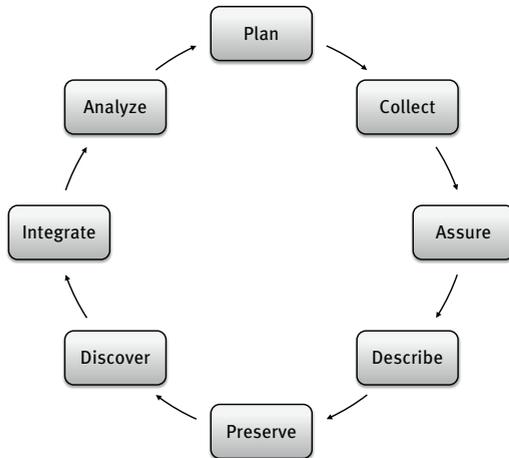


Abb. 2: Phasen des Lebenszyklus von Forschungsdaten aus der Perspektive des Forschers. Das hier abgebildete Schema wird im Projekt DataONE als Referenzmodell des Datenlebenszyklus verwendet (Lit. 16).

D 8.1 Daten entdecken

Entscheidend für den Zugang zu Forschungsdaten mit dem Ziel der Nachnutzung ist die Definition des Nutzungsszenarios und der Nutzergruppe (Lit. 19). In Bezug auf die Nachnutzung von Forschungsdaten können vier grundsätzliche Nutzungsszenarien unterschieden werden. Sie orientieren sich am primären Ziel der Datenhaltung (Dokumentation \leftrightarrow Nachnutzung) und der Zielgruppe (intern \leftrightarrow extern) (Lit. 20). Insbesondere in einem Szenario, in dem Daten extern nachgenutzt werden sollen, muss ein hoher Aufwand betrieben werden, diese Daten zu beschreiben und sie über Literatur, Datenportale und Kataloge findbar zu machen (Metadaten). Unterschieden wird hier zwischen Metadaten, die der Findbarkeit in Katalogen dienen (*discovery metadata*), und Metadaten mit detaillierten fachlichen Attributen zur tieferen Erschließung und Nachnutzung (*re-use metadata*). Für die reine Verbreitung von Metadaten in Katalogen und Portalen hat, auf Grund seiner Robustheit und inhaltlichen Flexibilität, das OAI-PMH Protokoll inzwischen eine weite Verbreitung gefunden (Lit. 21, Lit. 22).

Eine wichtige Komponente für die Nachnutzung von Daten ist der Einsatz von persistenten Identifikatoren, denn sie erlauben es, einen Datensatz eindeutig zu identifizieren und, wenn die Identifikatoren mit einem Auflösungsdienst verbunden sind, den Datensatz im Internet dauerhaft zu finden (Lit. 23). Diese eindeutige Referenzierung ermöglicht es auch, Datensätze analog und als Ergänzung zur wissenschaftlichen Literatur zu zitieren (Lit. 24, Lit. 25).

Mit der Vergabe von Identifikatoren für Forschungsdaten und der Sicherstellung ihrer langfristigen Auflösbarkeit kann hier auch ein neues Aufgabenfeld für akademische Bibliotheken entstehen, die als Dienstleister und als Gedächtnisorganisation die Veröffentlichung und die dauerhafte Verfügbarkeit von Daten (deren Langzeitarchivierung) unterstützen können (Lit. 13, Lit. 26). Diese neuen Aufgaben werden auch eine Weiterentwicklung der Qualifikationen der Mitarbeiter in den Bibliotheken erfordern und möglicherweise neue Berufsfelder an der Schnittstelle zwischen Datenproduzenten und Bibliotheken hervorbringen (Lit. 27).

Eng verbunden mit der Findbarkeit von Daten ist deren Verfügbarkeit. Forschungsdaten sind Teil der wissenschaftlichen Überlieferung, komplementär zur Literatur, in der sie interpretiert werden. Entsprechend forderte die „Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen“, dass analog zur wissenschaftlichen Literatur auch die Daten, die Grundlage einer Veröffentlichung waren, zugänglich gemacht werden sollen (Lit. 28). Mit dem Ziel, Daten aus öffentlich geförderter Forschung zugänglich zu machen, veröffentlichte 2006 die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) eine entsprechende Empfehlung an ihre Mitglieder (Lit. 29). Diese schloss den Auftrag an ihre Mitglieder ein, diese Empfehlung in nationale Gesetzgebung zu überführen.

Dieser *von oben verordnete* Ansatz steht im Gegensatz zu anderen Open Data Initiativen (siehe C 2 Open Access/Open Content, C 7 Web Science), da die Motive der Akteure in Open Data Initiativen und bei Open Access zu Forschungsdaten sehr verschieden liegen (Lit. 30, Lit. 31).

D 8.2 Daten erhalten

Digitale Forschungsdaten unterscheiden sich nicht grundsätzlich von anderen digitalen Objekten. Kennzeichnend für den Umgang mit Forschungsdaten ist die große Vielfalt der Formate und Strukturen, in denen Forschungsdaten entstehen (Lit. 32). Die beiden wichtigsten Hürden, die einer Nachnutzung der Daten im Weg stehen, sind hier der Verlust des Kontexts und das Veralten eines Dateiformats. Beide Probleme sind miteinander verbunden, denn mit dem Verlust des Kontexts geht auch oft das Wissen über die Entstehung, die Verwendung eines Datensatzes und seines Formats verloren. Auch hier helfen persistente Identifikatoren und Techniken des Semantic Web (siehe B 7 Semantic Web und Linked Open Data), diesen Kontext zu erhalten (Lit. 33).

Das Bewahren der Kontext- und Nutzungsinformationen wiederum basiert auf einem Bitstrom, dem eigentlichen digitalen Objekt. Diesen Komplex aus digitalem Objekt und seinen zugehörigen Kontext- und Nutzungsinformationen zu erhalten, ist die Aufgabe von Archivsystemen, deren Systematik z. B. im Open Archival Information Systems Referenzmodell (OAIS) beschrieben ist (Lit. 34). Ausgehend von diesem Referenzmodell sind in den vergangenen Jahren viele weitergehende Arbeiten entstanden. Für die Integrität der wissenschaftlichen Überlieferung ist dabei entscheidend, dass die Daten, wie ja auch in der „Berliner Erklärung“ gefordert, in einem vertrauenswürdigen Archiv vorgehalten werden (Lit. 35; vgl. D 4 Langzeitarchivierung).

D 8.3 Daten beschreiben

Daten ohne Beschreibung sind für eine Nachnutzung weitgehend unbrauchbar. Dieser Mangel wird umso deutlicher, je weiter (räumlich, zeitlich, fachlich) der Datenkonsument vom Kontext der Entstehung der Daten entfernt ist. Metadaten, die formale Beschreibung eines Datensatzes, sind ein Thema das immer wieder hitzige Debatten heraufbeschwören kann, denn es gibt viele Wege, einen Datensatz mit Attributen zu versehen (vgl. B 2 Metadaten). Wie bereits erwähnt, muss hier grundsätzlich unterschieden werden zwischen Metadaten, die der Findbarkeit in Katalogen und Portalen dienen, und Metadaten mit detaillierten fachlichen Attributen, mit denen die Nachnutzbarkeit der Daten unterstützt wird.

Bei Metadaten, die der Findbarkeit von Daten in Katalogen und Portalen dienen, reicht in vielen Fällen bereits ein einfaches bibliographisches Schema, wie zum Beispiel Dublin Core oder DataCite (Lit. 36). Teilweise haben sich bereits Standards etabliert, wie zum Beispiel ISO 19115 für georeferenzierte Daten (Lit. 37). Möglich ist dabei auch, einen Datensatz mit verschiedenen Metadatenprofilen zu beschreiben, denn je nach vorgesehener Anwendung können sehr unterschiedliche Metadaten-Elemente zur Beschreibung relevant sein (Lit. 38). Dabei fällt auf, dass neben einer syntaktischen auch eine semantische Interoperabilität notwendig ist (z. B. Lit. 39, Lit.

40). Für deren Möglichkeiten und Grenzen verweisen wir auf B 7 Semantic Web und Linked Open Data.

Nachnutzung von Forschungsdaten setzt Vertrauen in die Quellen voraus. Während es seit langem Teil der Ausbildung ist, die Glaubwürdigkeit von Literatur einschätzen zu lernen, herrscht in Bezug auf die Qualität von Forschungsdaten noch Unsicherheit (Lit. 41). Vertrauenswürdigkeit bedeutet hier nicht Rechtssicherheit, sondern ist vielmehr eine subjektive Einschätzung der Nutzbarkeit im Sinne der Definition von Qualität im Standard ISO 9000 (Lit. 42, Lit. 43). Dies beinhaltet aber auch Vertrauen in den Entstehungsprozess der Daten (Lit. 43). Zusammen mit der Vielfalt, mit der Daten strukturiert und beschrieben werden können, gibt es eine ebenso große Zahl an Möglichkeiten, die Qualität dieser Daten und ihrer Metadaten zu beschreiben. Vereinzelt gibt es bereits Beschreibungsmetriken für die Qualität von Forschungsdaten. Es besteht jedoch noch Forschungsbedarf.

D 8.4 Forschungsdaten verwalten

Der aufwendigste und teuerste Schritt ist, Forschungsdaten in eine Forschungsdateninfrastruktur zu überführen. In Fällen, in denen Daten automatisiert entstehen, zum Beispiel in Messnetzen oder in Simulationen, ist es relativ leicht, die Metadaten im Entstehungsprozess der Daten gleich mit zu erzeugen. Sehr viel schwieriger ist es, die große Anzahl der Datensätze des *small data* zu erfassen (Lit. 04). Zwar stellt das Datenvolumen keine Herausforderung dar, jedoch sind die Datenstrukturen heterogen, die Metadaten können nicht ohne weiteres automatisch erzeugt werden.

Gerade an der Schnittstelle zwischen dem Arbeitsbereich eines einzelnen Wissenschaftlers oder einer Forschergruppe und einer Forschungsdateninfrastruktur wird deutlich sichtbar, dass die Anforderungen vielseitiger kaum sein können. Treloar u. a. (Lit. 44) nennen dies das *data curation continuum*. Ein Kontinuum lässt sich jedoch organisatorisch nicht handhaben, deswegen schlagen Treloar u. a. vor, dieses Kontinuum in Domänen zu gliedern, in denen jeweils andere Akteure die Verantwortung für die Daten übernehmen. Die Anzahl der Domänen richtet sich nach den örtlichen Gegebenheiten, eine Aufteilung in vier Domänen (siehe Abbildung 3) erwies sich bisher als ein Konzept, das den Gegebenheiten in vielen Forschungseinrichtungen gerecht wird.

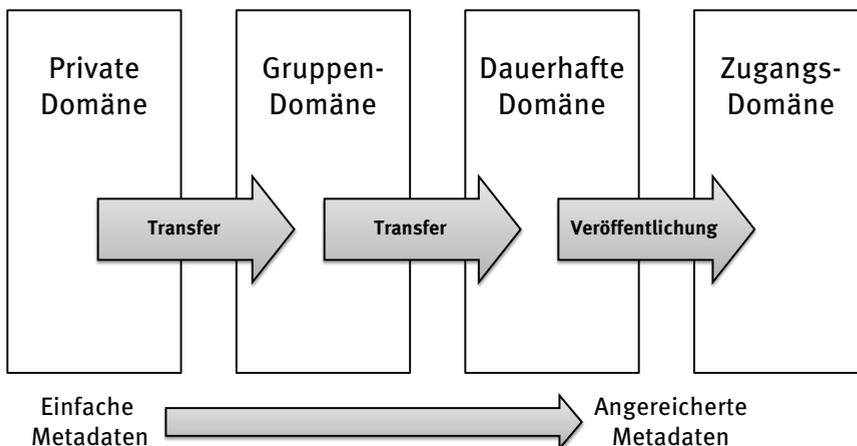


Abb. 3: Vorschlag zur Gliederung des Data Curation Continuum nach Treloar et al. (Lit. 44) in Zuständigkeitsdomänen (Lit. 45)

Zu diesem Zeitpunkt ist es auch bereits angebracht, eine Richtungsentscheidung zur Lizenzierung von Forschungsdaten zu fällen (vgl. C 8 Lizenzierungsformen), denn eine spätere Änderung der Nutzungsbedingungen kann sehr aufwendig sein. Die Wahl einer ungeeigneten – meist zu restriktiven – Lizenz kann die Nutzbarkeit von Forschungsdaten erheblich erschweren (Lit. 46, Lit. 47, Lit. 48, Lit. 49). Leider ist der urheberrechtliche Status von Forschungsdaten noch unklar (Lit. 50). Zudem ist der Prestigegewinn, den ein Wissenschaftler durch sorgfältige Verwaltung seiner Daten erzielen kann, nur gering im Vergleich zu Veröffentlichungen. Folglich ist, ungeachtet aller bestehenden Vorgaben, seine Motivation zur systematischen Verwaltung seiner Daten gering (Lit. 31, Lit. 51).

D 8.5 Umgang mit Forschungsdaten planen

Als Konsequenz aus einer Reihe von Wissenschaftsskandalen in den 1990er Jahren veröffentlichte die DFG „Regeln für eine gute wissenschaftliche Praxis“ (Lit. 52), die in weitgehend ähnlichen Formen von den anderen deutschen Wissenschaftsorganisationen und den Hochschulen übernommen wurden. In Abschnitt 7 „Primärdaten“ verlangt die DFG die Sicherung von Forschungsdaten:

Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden. (Lit. 52)

Das Ziel dieser Empfehlung war jedoch nur die Beweissicherung, die Nutzbarkeit dieser Daten wurde nicht thematisiert.

Wie bereits erwähnt, veröffentlichte die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) 2006 eine Empfehlung an ihre Mitglieder über den Umgang mit Daten aus öffentlich geförderter Forschung (Lit. 29). In der Folge modifizierte die DFG ihre Vergaberichtlinien und ihren Leitfaden für Antragsteller (Lit. 53). Seit 2011 werden Antragsteller aufgefordert, einen Plan vorzulegen, welche Daten im Projekt erhoben werden und wie diese Daten für eine Nutzung bereitgestellt werden sollen. Da es sich noch um eine neue Entwicklung handelt, liegen noch keine gesicherten Erfahrungen aus der Praxis vor. Es kann jedoch davon ausgegangen werden, dass es hier noch lange deutliche Unterschiede in der jeweiligen Praxis der Wissenschaftsdisziplinen geben wird (Lit. 54, Lit. 55).

D 8.6 Datenintegration und Analyse – Daten-getriebene Forschung

Bei der Datenintegration und Analyse überschneiden sich die oben geschilderten Aspekte des Umgangs mit Forschungsdaten, denn Integration und Analyse größerer Datenbestände ist nur möglich, wenn die Daten auffindbar sind und auch zur Verfügung stehen. Anhand der Metadaten, die im Verlauf des Datenlebenszyklus immer weiter angereichert werden, kann der Forscher eine Entscheidung treffen, ob die gefundenen Daten für ihn brauchbar sind. Metadaten können auch Auskunft über die Qualität und Vertrauenswürdigkeit der Daten geben.

Der wissenschaftliche Mehrwert ergibt sich jedoch nicht aus Datenobjekten und dazugehörigen Metadaten, sondern aus deren Interpretation. Die Entwicklung der Informationstechnologie erlaubt es heute, auch große Datenmengen explorativ zu erforschen und aus den erkannten Mustern immer neue Hypothesen an den Daten zu formulieren und zu überprüfen (Lit. 03). Solche datengetriebenen Ansätze werden zunehmend angewandt, um komplexe Systeme und Prozesse besser zu verstehen. Grundidee der datengetriebenen Forschung ist, aus vorhandenen Daten möglichst umfassend all die Information zu extrahieren, die für die Lösung einer bestimmten Fragestellung hilfreich ist. Mit Hilfe eines umfangreichen Methodenrepertoires sollen Muster und Zusammenhänge in den Daten erkannt und extrahiert werden, um daraus Hypothesen zu generieren oder zu

verifizieren. „Let the data speak“ ist der Grundgedanke des Daten-getriebenen Ansatzes. Ob es sich hier um ein neues Paradigma in der Forschung handelt (Lit. 02, Lit. 03) oder ob die Informationstechnologie nur neue und sehr leistungsfähige Werkzeuge bereitstellt, die den bisherigen Methodenansatz erweitern, wird derzeit noch debattiert. Absehbar ist allerdings, dass Wissenschaft ohne Anwendung solcher Werkzeuge sowohl auf Texte als auch auf Forschungsdaten mittelfristig allein aufgrund der Menge der zu verarbeitenden Quellen vor großen Herausforderungen steht (vgl. B 4 Text Mining und Data Mining).

D 8.7 Zusammenfassung und Ausblick

Auffallend beim Umgang mit Forschungsdaten ist deren große Vielfalt (Lit. 32). Die Entwicklung des Internets und der Informationstechnologie in den vergangenen zwanzig Jahren haben dazu geführt, dass in ungeahntem Maß Forschungsdaten erhoben, produziert und ausgetauscht werden (Lit. 01, Lit. 02). Ein Buchkapitel über den Umgang mit Forschungsdaten kann das Thema daher nur kurz beleuchten und auf die weiterführende Literatur verweisen (z. B. Lit. 09, Lit. 17, Lit. 18). Bisher war der Umgang mit Forschungsdaten dort am weitesten entwickelt, wo in kollaborativen Strukturen gearbeitet wurde (Lit. 32). Die hier stattfindende Entwicklung von Virtuellen Forschungsumgebungen (VRE) wird sich mit der Zeit auch in andere Forschungsgebiete ausbreiten. Die Datenmanagement-Werkzeuge der Zukunft werden allerdings nicht mehr die monolithischen Anwendungen von heute sein, sondern modular auf Dateninfrastrukturen aufbauen (Lit. 38, Lit. 51). Eine Modularisierung der Datenmanagement-Werkzeuge – oder auch „Appification“ aus der Sicht der Forscher – wird zu einer engeren Verzahnung des Datenmanagements mit der wissenschaftlichen Arbeit führen und damit den Umgang mit Forschungsdaten merklich erleichtern und verbessern.

Gleichzeitig mit der zunehmenden Verbreitung und Nachnutzung von Forschungsdaten nimmt auch das Bewusstsein für den Wert von Forschungsdaten als Grundlage der Wissenschaft zu. Forschungsförderer verlangen von ihren Zuwendungsempfängern Datenmanagementpläne, auch Fachzeitschriften haben den Wert von Daten erkannt und sehen zugängliche Forschungsdaten als Mehrwert zu den veröffentlichten Artikeln. Schon jetzt zeigt sich, dass Artikel, deren Daten frei zugänglich sind, häufiger und über einen längeren Zeitraum zitiert werden als Artikel bei denen die Daten nicht zugänglich sind (Lit. 56, Lit. 57, Lit. 58, Lit. 59). Erst die Integration des Forschungsdatenmanagements in das Wertesystem der Wissenschaft, verbunden mit attraktiven technischen und organisatorischen Angeboten, wird dazu führen, dass der Umgang mit Forschungsdaten verbessert wird, so dass diese auch langfristig zur Nachnutzung zur Verfügung stehen.

Schon heute wird sichtbar, welches Potenzial in der Synthese und Neuanalyse von bereits vorhandenen Daten steckt. Zusätzlich wird es einfacher werden, veröffentlichte Forschungsergebnisse nachzuvollziehen, wodurch sowohl der Erkenntnisgewinn als auch die Integrität der Forschung an Qualität gewinnen werden. Diese Entwicklung vollzieht sich heute noch langsam, wird sich aber mit der Zeit durch eine Verbesserung der Rahmenbedingungen weiter beschleunigen.

Literatur

- 01 Hey, T.; Trefethen, A.: e-Science and its implications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 361, 1809-1825 (doi:10.1098/rsta.2003.1224)
- 02 Szalay, A; Gray, J: 2020 Computing: Science in an exponential world. *Nature* 440, 413-414, 2006 (doi:10.1038/440413a)
- 03 Hey, T; Tansley, S; Tolle, K (editors): *The Fourth Paradigm: Data-Intensive Scientific Discovery*. 1.1 Ausgabe. Redmond, WA: Microsoft Research. 287 S., 2009 (<http://research.microsoft.com/en-us/collaboration/fourthparadigm/>)

- 04 Onsrud, HJ; Campbell, J: Big opportunities in access to 'small science' data. *DSJ* 6: OD58-OD66, 2007 (doi:10.2481/dsj.6.OD58)
- 05 McNally, R; Mackenzie, A; Hui, A; Tomomitsu, J: Understanding the 'Intensive' in 'Data Intensive Research': Data Flows in Next Generation Sequencing and Environmental Networked Sensors. *IJDC* 7: 81-94, 2012 (doi:10.2218/ijdc.v7i1.216)
- 06 Baggerly, K: Disclose all data in publications. *Nature* 467: 401, 2010 (doi:10.1038/467401b)
- 07 Hanson, B; Sugden, A; Alberts, B: Making Data Maximally Available. *Science* 331: 649, 2011 (doi:10.1126/science.1203354)
- 08 Sommer, J: The delay in sharing research data is costing lives. *Nature Medicine* 16: 744, 2010 (doi:10.1038/nm0710-744)
- 09 Royal Society: Science as an open enterprise. London, Großbritannien: The Royal Society, 2012 (<http://royalsociety.org/policy/projects/science-public-enterprise/report/>)
- 10 Allianzinitiative Digitale Information: Grundsätze zum Umgang mit Forschungsdaten. Bonn: Allianz der deutschen Wissenschaftsorganisationen, 2010 (<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>)
- 11 DFG: Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. Bonn: Deutsche Forschungsgemeinschaft 2009 (http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/veroeffentlichungen/dokumentationen/download/ua_inf_empfehlungen_200901.pdf)
- 12 Mervis, J: NSF to Ask Every Grant Applicant for Data Management Plan – ScienceInsider. *Science Insider*. 2010 (<http://news.sciencemag.org/scienceinsider/2010/05/nsf-to-ask-every-grant-applicant.html>)
- 13 Klump, J; Bertelmann, R; Brase, J; Diepenbroek, M; Grobe, H et al.: Data publication in the Open Access Initiative. *Data Science Journal* 5: 79-83, 2006 (doi:10.2481/dsj.5.79)
- 14 Staudigel, H; Helly, J; Koppers, A; Shaw, H; McDonough, WF et al.: Electronic data publication in geochemistry. *Geochemistry, Geophysics, Geosystems – G (super 3)* 4: 17, 2003 (doi:10.1029/2002GC000314)
- 15 Klump, J; Bertelmann, R: Mehr als nur die Emulation von Papier. *eScience & Forschungsdatenmanagement*. Potsdam: Fachhochschule Potsdam, 2010 (<http://informationswissenschaften.fh-potsdam.de/abstracts.html#c13827>)
- 16 Strasser, C; Cook, R; Michener, W; Budden, A: DataONE Primer on Data Management. Albuquerque, NM: University of New Mexico, 2012 (http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf)
- 17 Neuroth, H; Strathmann, S; Oßwald, A; Scheffel, R; Klump, J et al. (Hrsg.): Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme. Boizenburg: Verlag Werner Hülsbusch. 378 S., 2012 (<http://nestor.sub.uni-goettingen.de/bestandsaufnahme>)
- 18 Büttner, S; Hobohm; H-C, Müller; L (editors): *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock + Herrchen, 2011 (<http://opus.kobv.de/fhpotdams/volltexte/2011/241/pdf/HandbuchForschungsdatenmanagement.pdf>)
- 19 Parsons, MA; Duerr, R: Designating user communities for scientific data: challenges and solutions. *Data Science Journal* 4: 31-38, 2005 (doi:10.2481/dsj.4.31)
- 20 Ludwig, J: Infrastruktureinrichtungen und Forschungsdaten: Sichtweise und Service, 2012 (<http://www.tmf-ev.de/DesktopModules/Bring2mind/DMX/Download.aspx?EntryId=17025&PortalId=0>)
- 21 Devarakonda, R; Palanisamy, G; Green, J; Wilson, B: Data sharing and retrieval using OAI-PMH. *Earth Science Informatics*: 1-5, 2010 (doi:10.1007/s12145-010-0073-0)
- 22 Van de Sompel, H; Nelson, ML; Lagoze, C; Warner, S: Resource Harvesting within the OAI-PMH Framework. *D-Lib Magazine* 10: 18, 2004 (doi:10.1045/december2004-vandesompel)
- 23 Simons, N: Implementing DOIs for Research Data. *D-Lib* 18. 2012 (doi:10.1045/may2012-simons)
- 24 Ball, A; Duke, M: *How to Cite Datasets and Link to Publications*. Edinburgh, Großbritannien: Digital Curation Centre, 2012 (<http://www.dcc.ac.uk/resources/how-guides/cite-datasets>)
- 25 Repository evaluation, selection, and coverage policies for the Data Citation Index within Thomson Reuters Web of Knowledge (2012). Philadelphia, PA: Thomson Reuters (<http://wokinfo.com/media/pdf/>)

- DCI_selection_essay.pdf)
- 26 Smith, M: Communicating with data: new roles for scientists, publishers and librarians. *Learned Publishing* 24: 203-205, 2011 (doi:10.1087/20110308)
 - 27 Auckland, M: Re-skilling for Research. London, Großbritannien: Research Libraries UK, 2012 (<http://www.rluk.ac.uk/files/RLUK%20Re-skilling.pdf>)
 - 28 Berlin Declaration: Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003 (<http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>)
 - 29 OECD: Recommendation of the Council concerning Access to Research Data from Public Funding. Paris, Frankreich: Organisation for Economic Co-operation and Development, 2006
 - 30 Klump, J: Offener Zugang zu Forschungsdaten: Open Data und Open Access to Data – Die ungleichen Geschwister. Herb, U (Hrsg.), *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft*. Saarbrücken: universaar, 45-53, 2012 (<http://nbn-resolving.de/urn:nbn:de:bsz:291-universaar-873>)
 - 31 Borgman, CL: The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology* 63: 1059-1078, 2012 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155)
 - 32 Ludwig, J: Zusammenfassung. Neuroth, H; Strathmann, S; Oßwald, A; Scheffel, R; Klump, J et al. (Hrsg.), *Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme*. Boizenburg: Verlag Werner Hülsbusch. 295-310, 2012 (http://www.nestor.sub.uni-goettingen.de/bestandsaufnahme/kapitel/nestor_bestandsaufnahme_015.pdf)
 - 33 Pepe, A; Mayernik, M; Borgman, CL; Van de Sompel, H: From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. *JASIST*: 28, 2009 (<http://arxiv.org/abs/0906.2549v3>)
 - 34 CCSDS: Reference Model for an Open Archival Information System (OAIS). Magenta Book. Recommended Practice. Greenbelt, MD: Consultative Committee for Space Data Systems, 2012 (<http://public.ccsds.org/publications/archive/650x0m2.pdf>)
 - 35 Klump, J: Criteria for the Trustworthiness of Data Centres. *D-Lib Magazine* 17, 2011 (doi:10.1045/january2011-klump)
 - 36 Starr, J; Gastl, A: isCitedBy: A Metadata Scheme for DataCite. *D-Lib Magazine* 17, 2011 (doi:10.1045/january2011-starr)
 - 37 Shaon, A; Woolf, A: Long-term Preservation for Spatial Data Infrastructures: a Metadata Framework and Geo-portal Implementation. *D-Lib Magazine* 17, 2011 (doi:10.1045/september2011-shaon)
 - 38 Razum, M; Schwichtenberg, F; Wagner, S; Hoppe, M: eSciDoc Infrastructure: A Fedora-Based e-Research Framework. *Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science*. Heidelberg: Springer Verlag, Bd. 5714, 227-238. 2009 (doi:10.1007/978-3-642-04346-8_23)
 - 39 Hughes, S; Crichton, D; Kelly, S; Mattmann, CA; Crichton, J et al.: Intelligent resource discovery using ontology-based resource profiles. *Data Science Journal* 4: 171-188, 2005 (doi:10.2481/dsj.4.171)
 - 40 Hall, W; De Roure, D; Shadbolt, N: The evolution of the Web and implications for eResearch. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367: 991-1001, 2009 (doi:10.1098/rsta.2008.0252)
 - 41 Kaiser, J: Data Integrity Report Sends Journals Back to the Drawing Board. *Science* 325: 381, 2009 (doi:10.1126/science.325_381)
 - 42 ISO 9000:2000: Quality management systems – Fundamentals and vocabulary. Standard. Genf, Schweiz: International Organization for Standardization (ISO). (http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=29280)
 - 43 Lee, YW; Strong, DM; Kahn, BK; Wang, RY: AIMQ: a methodology for information quality assessment. *Information & Management* 40: 133-146, 2002 (doi:10.1016/S0378-7206(02)00043-5)
 - 44 Treloar, A; Groenewegen, D; Harboe-Ree, C: The Data Curation Continuum – Managing Data Objects in Institutional Repositories. *D-Lib Magazine* 13: 13 (doi:10.1045/september2007-treloar)
 - 45 Klump, J: Langzeiterhaltung digitaler Forschungsdaten. Büttner, S; Hobohm, H-C; Müller, L (Hrsg.), *Handbuch Forschungsdatenmanagement*. Bad Honnef, Germany: Bock + Herrchen, 115-122, 2011 (<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:kobv:525-opus-2339>)
 - 46 Klimpel, P: Folgen, Risiken und Nebenwirkungen der Bedingung „nicht-kommerziell – NC“. Berlin:

- Wikimedia Deutschland, 2012 (http://irights.info/userfiles/CC-NC_Leitfaden_web.pdf)
- 47 Ball, A: How to License Research Data. Edinburgh, Großbritannien: Digital Curation Centre, 2011 (<http://www.dcc.ac.uk/resources/how-guides/license-research-data>)
- 48 Rees, J: Recommendations for independent scholarly publication of data sets. San Francisco, CA: Creative Commons, 2010 (<http://neurocommons.org/report/data-publication.pdf>)
- 49 Hrynaszkiewicz, I; Cockerill, MJ: Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journals. *BMC Research Notes* 5: 494, 2012 (doi:10.1186/1756-0500-5-494)
- 50 de Cock Buning, M; van Dinther, B; Jeppersen de Boer, C, Ringnalda, A: The legal status of research data in the Knowledge Exchange partner countries. Kopenhagen, Dänemark: Knowledge Exchange, 2011 (<http://www.knowledge-exchange.info/Default.aspx?ID=461>)
- 51 Feijen, M: What researchers want. Utrecht, The Netherlands: SURFfoundation, 2011 (<http://www.surfoundation.nl/en/publicaties/Pages/Whatresearcherswant.aspx>)
- 52 DFG: Sicherung guter wissenschaftlicher Praxis. Bonn: Deutsche Forschungsgemeinschaft, 1998 (http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf)
- 53 DFG: Merkblatt für Anträge auf Sachbeihilfen mit Leitfaden für die Antragstellung und ergänzenden Leitfäden für die Antragstellung für Projekte mit Wertungspotenzial, für die Antragstellung für Projekte im Rahmen einer Kooperation mit Entwicklungsländern. Bonn: Deutsche Forschungsgemeinschaft (DFG). 2010 (http://www.dfg.de/download/formulare/1_02/1_02.pdf)
- 54 Tenopir, C; Allard, S; Douglass, K; Aydinoglu, AU; Wu, L et al.: Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* 6: e21101 (doi:10.1371/journal.pone.0021101)
- 55 Piwowar, HA: Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. *PLoS ONE* 6: e18657. 2011 (doi:10.1371/journal.pone.0018657)
- 56 Sears, JR: Data Sharing Effect on Article Citation Rate in Paleocyanography. *EOS, Transactions, American Geophysical Union* 92: IN53B-1628, 2011 (<http://adsabs.harvard.edu/abs/2011AGUFMIN53B1628S>)
- 57 Henneken, EA; Accomazzi, A: Linking to Data – Effect on Citation Rates in Astronomy. *arXiv:1111.3618*, 2011 (<http://arxiv.org/abs/1111.3618>)
- 58 Botstein, D: It's the Data! *Molecular Biology of the Cell* 21: 4-6, 2010 (doi:10.1091/mbc.E09-07-0575)
- 59 Piwowar, HA; Day, RS; Fridsma, DB: Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2: e308, 2007 (doi:10.1371/journal.pone.0000308)