

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.0

Kapitel 17.10
Digitale Forschungsdaten

nestor 

vwth

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn
im Rahmen des Projektes: nestor – Kompetenznetzwerk Langzeitarchivierung und
Langzeitverfügbarkeit digitaler Ressourcen für Deutschland
nestor – Network of Expertise in Long-Term Storage of Digital Resources
<http://www.langzeitarchivierung.de/>

Kontakt: editors@langzeitarchivierung.de
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Die Herausgeber danken Anke Herr (Korrektur), Martina Kerzel (Bildbearbeitung) und
Jörn Tietgen (Layout und Formatierung des Gesamttextes) für ihre unverzichtbare
Unterstützung bei der Fertigstellung des Handbuchs.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter
<http://www.d-nb.de/> abrufbar.

Die Inhalte dieses Buchs stehen auch als Onlineversion
(<http://nestor.sub.uni-goettingen.de/handbuch/>)
sowie über den Göttinger Universitätskatalog (<http://www.sub.uni-goettingen.de>) zur
Verfügung.

Die digitale Version 2.0 steht unter folgender Creative-Commons-Lizenz:
„Attribution-Noncommercial-Share Alike 3.0 Unported“
<http://creativecommons.org/licenses/by-nc-sa/3.0/>



Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch, Boizenburg.
© Verlag Werner Hülsbusch, Boizenburg, 2009
www.vwh-verlag.de
In Kooperation mit dem Universitätsverlag Göttingen

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und
als solche den gesetzlichen Bestimmungen unterliegen.

Druck und Bindung: Kunsthaus Schwanheide

Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-940317-48-3

URL für Kapitel 17.10 „Digitale Forschungsdaten“ (Version 2.0): [urn:nbn:de:0008-20090811921](http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20090811921)
<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20090811921>



Gewidmet der Erinnerung an Hans Liegmann (†), der als Mitinitiator und früherer Herausgeber des Handbuchs ganz wesentlich an dessen Entstehung beteiligt war.

17.10 Digitale Forschungsdaten

Jens Klump

Einführung

Durch eine Reihe von Aufsehen erregenden Wissenschaftsskandalen in den neunziger Jahren des 20. Jahrhunderts sah sich die Deutsche Forschungsgemeinschaft (DFG) gezwungen, „Empfehlungen für eine gute wissenschaftliche Praxis“ auszusprechen, die in vergleichbarer Form auch von anderen Wissenschaftsorganisationen übernommen wurden. In ihren Empfehlungen bezieht sich die DFG auf Daten, die Grundlage einer wissenschaftlichen Veröffentlichung waren. Sie verlangt von ihren Zuwendungsempfängern, dass diese Daten für mindestens zehn Jahre auf geeigneten Datenträgern sicher aufbewahrt werden müssen.⁹⁴ Für die einzelnen Disziplinen ist der Umgang mit Daten im einzelnen zu klären, um eine angemessene Lösung zu finden.⁹⁵ Diese Policy dient jedoch in erster Linie einer Art Beweissicherung; über Zugang zu den Daten und ihre Nutzbarkeit sagen die Empfehlungen nichts aus. Zudem ist bisher noch kein Fall bekannt geworden, in dem die DFG Sanktionen verhängt hätte, allein weil der Pflicht zur Archivierung von Daten nicht nachgekommen wurde.

Auf Einladung der DFG wurde im Januar 2008 in einem Rundgespräch der Umgang mit Forschungsdaten und deren Bereitstellung diskutiert. Auch die Auswirkungen der „Empfehlungen“ von 1998 wurden einem kritischen Review unterzogen.⁹⁶ Als Ergebnis dieser Konsultationen und der darauf folgenden Entscheidungen der Fachausschüsse veröffentlichte die DFG im Januar 2009 ihre „Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten“.⁹⁷

Welche Daten sollen und müssen archiviert werden? Die Empfehlungen der DFG sprechen von Forschungsprimärdaten. Dieser Begriff „Primärdaten“ sorgt immer wieder für Diskussion, denn die Definition des Begriffs ist sehr von der eigenen Rolle in der wissenschaftlichen Wertschöpfungskette bestimmt. Für den einen sind „Primärdaten“ der Datenstrom aus einem Gerät, z.B. einem Satelliten. In der Fernerkundung werden diese Daten „Level 0“

94 DFG (1998), Empfehlung 7

95 DFG (1998), Empfehlung 1

96 Kluttig (2008)

97 DFG (2009)

Produkte genannt. Für andere sind „Primärdaten“ zur Nachnutzung aufbereitete Daten, ohne weiterführende Prozessierungsschritte. Wieder andere differenzieren nicht nach Grad der Verarbeitung sondern betrachten alle Daten, die Grundlage einer wissenschaftlichen Veröffentlichung waren, als Primärdaten. Der begrifflichen Klarheit wegen sollte daher das Präfix „Primär-“ nicht mehr verwendet werden und statt dessen nur noch von wissenschaftlichen Daten oder Forschungsdaten gesprochen werden.

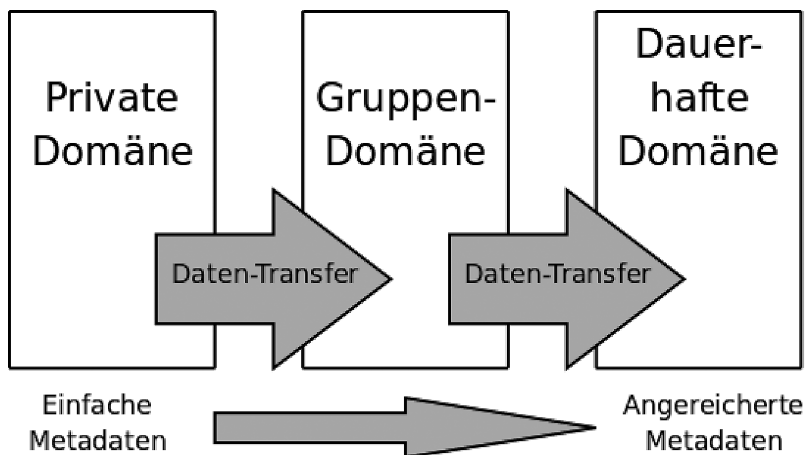
Welche Definition des Begriffs man auch wählt, wissenschaftliche Daten sind geprägt durch ihre Herkunft aus experimentellem Vorgehen, d.h. anders als Daten aus Arbeitsabläufen der Industrie oder Verwaltung stammen Forschungsdaten überwiegend aus informellen Arbeitsabläufen, die immer wieder ad hoc an die untersuchte Fragestellung angepasst werden.⁹⁸ Da in den meisten Fällen keine Formatvorgaben vorhanden sind, werden Forschungsdaten in einer Vielfalt von Dateiformaten hergestellt, die semantisch selten einheitlich strukturiert und nur lückenhaft mit Metadaten beschrieben sind. Diese Faktoren stellen für die digitale Langzeitarchivierung von Forschungsdaten eine größere Herausforderung dar, als die Datenmenge, auch wenn diese in einzelnen Fällen sehr groß sein kann.⁹⁹

Im Laufe des digitalen Lebenszyklus von Forschungsdaten werden zudem in den verschiedenen Phasen sehr unterschiedliche Anforderungen an die Persistenz der Daten und der Werkzeuge zum Umgang mit Forschungsdaten gestellt. Zwischen dem Entstehen der Daten in wissenschaftlichen Arbeitsprozessen und der sicheren, nachnutzbaren Archivierung der Daten besteht ein breites Spektrum von teilweise gegensätzlichen Anforderungen, auch *Digital Curation Continuum* genannt. Organisatorisch ist ein Kontinuum allerdings nicht handhabbar, weswegen es notwendig ist, innerhalb einer Organisation zu bestimmen, wer in welcher Phase des Lebenszyklus von Forschungsdaten für deren Pflege verantwortlich ist. Auf Grund des vorhandenen Kontextwissens reicht in den Phasen vor der Speicherung in der dauerhaften Domäne ein eingeschränktes Metadatenprofil aus, das bei der Überführung in die nächste Domäne (teil-)automatisch angereichert werden kann.¹⁰⁰

98 Barga & Gannon (2007)

99 Klump (2008)

100 Treloar & Harboe-Ree (2008); Treloar et al. (2007)



Für den Forscher liegt es nicht im Fokus seines wissenschaftlichen Arbeitens, Daten zu archivieren und zugänglich zu machen, denn bisher bestehen keine Anreize an Wissenschaftler, zumindest Daten, die Grundlage einer Veröffentlichung waren, für andere zugänglich zu machen.¹⁰¹ Nur an sehr wenigen Stellen besteht heute im wissenschaftlichen Veröffentlichungssystem oder in der Forschungsförderung die Pflicht, Forschungsdaten für andere zugänglich zu machen. Darüber hinaus ist nicht geklärt, wer für die Langzeitarchivierung von Forschungsdaten verantwortlich ist und wie diese Leistung finanziert wird.¹⁰² Dies führt zu Defiziten im Management und in der Archivierung wissenschaftlicher Daten mit möglichen negativen Folgen für die Qualität der Forschung.¹⁰³ Diese Entwicklung ist inzwischen auch den Herausgebern wissenschaftlicher Fachzeitschriften bewusst geworden. Als Konsequenz daraus verlangen bereits einige Zeitschriften, dass die Daten, die Grundlage der eingereichten Arbeit sind, für die Gutachter und später auch für die Leser zugänglich sind.

Trotz der Empfehlungen für eine gute wissenschaftliche Praxis sind kohärente Datenmanagementstrategien, Archivierung von Forschungsdaten und, soweit möglich, Zugang zu Daten meist nur in größeren Forschungsverbänden zu finden, die für Erfolge in der Forschung auf enge Zusammenarbeit angewiesen sind, oder in Fällen, in denen es gesetzliche Vorgaben für den Umgang mit Daten gibt. Wie schon in der Diskussion um den Offenen Zugang zu wissen-

101 Klump et al. (2006)

102 Lyon (2007)

103 Nature Redaktion (2006)

schaflichem Wissen (Open Access) zeigt sich hier, dass eine Policy nur wirksam ist, wenn sie eine Verpflichtung mit sich bringt und gleichzeitig Anreize zur Zusammenarbeit bietet.¹⁰⁴ Um das Ziel einer nachhaltigen digitalen Langzeitarchivierung von Forschungsdaten zu erreichen, muss sowohl eine organisatorische Strategie verfolgt werden, die Langzeitarchivierung von Daten zu einem anerkannten Beitrag zur wissenschaftlichen Kultur macht und die gleichzeitig von einer technischen Strategie unterstützt wird, die den Akteuren für die digitale Langzeitarchivierung von wissenschaftlichen Forschungsdaten geeignete Werkzeuge in die Hand gibt. Mit dazu gehören eine Professionalisierung des Datenmanagements und der digitalen Langzeitarchivierung von Forschungsdaten auf Seiten der Projekte und Archive.

Organisatorische Strategien

Auf Grund der enormen Summen, die jährlich für die Erhebung wissenschaftlicher Daten ausgegeben werden, beschäftigt sich die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) bereits seit einigen Jahren mit der Frage, wie mit Daten aus öffentlich geförderter Forschung umgegangen werden sollte. Auf dem Treffen der Forschungsminister im Januar 2004 wurde beschlossen, dass der Zugang zu Daten aus öffentlich geförderter Forschung verbessert werden muss.¹⁰⁵ Mit diesem Mandat im Hintergrund befragte die OECD die Wissenschaftsorganisationen ihrer Mitgliedsländer zu deren Umgang mit Forschungsdaten. Basierend auf den Ergebnissen der Befragung wurde eine Studie verfasst und im Dezember 2006 verabschiedete der Rat der OECD eine „Empfehlung betreffend den Zugang zu Forschungsdaten aus öffentlicher Förderung“.¹⁰⁶ Diese Empfehlung ist bindend und muss von den Mitgliedsstaaten der OECD in nationale Gesetzgebung umgesetzt werden, die Umsetzung wird von der OECD beobachtet. In Abschnitt M der Empfehlung wird vorgeschlagen, dass schon bei der Planung von Projekten eine nachhaltige, langfristige Archivierung der Daten berücksichtigt wird.

Parallel dazu, und mit mehr Aufsehen in der Öffentlichkeit, wurde im Oktober 2003 von den Wissenschaftsorganisationen die „Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen“ veröffentlicht, deren Schwerpunkt auf dem Zugang zu wissenschaftlicher Literatur für Forschung und Lehre liegt.¹⁰⁷ In ihre Definition des offenen Zugangs bezieht die „Berliner

104 Bates et al. (2006); Spittler (1967)

105 OECD (2004)

106 OECD (2007)

107 Berliner Erklärung (2003)

Erklärung“ auch Daten und Metadaten mit ein. Die Langzeitarchivierung ist hier ein Mittel zum Zweck, das den offenen Zugang zu wissenschaftlichem Wissen über das Internet auf Dauer ermöglichen soll. Aufrufe dieser Art wurden stets begrüßt, aber blieben leider ohne Folge.¹⁰⁸ Dieses Problem betrifft die Institutional Repositories des Open Access genauso wie die Datenarchive. Es sollte daher geprüft werden, inwiefern die Strategien, die bei der Umsetzung von Open Access angewandt werden, sich auch auf den offenen Zugang zu Daten anwenden lassen.¹⁰⁹

Wenngleich es einige Policies gibt, die den Zugang zu Daten ermöglichen sollen, so hat sich erst recht spät die Erkenntnis durchgesetzt, dass die digitale Langzeitarchivierung von Forschungsdaten eine Grundvoraussetzung des offenen Zugangs ist. Eine umfangreiche Studie wurde dazu bereits in der ersten Förderphase des Projekts nestor erstellt.¹¹⁰ Eine ähnliche Studie wurde auch für das britische Joint Information Systems Committee (JISC) veröffentlicht¹¹¹ und das Thema in einer weiteren Studie vertieft¹¹². Einzelne Systeme, die als Best-Practice Beispiele gelten dürfen, da sie die Voraussetzungen von Offenem Zugang und vertrauenswürdiger digitaler Langzeitarchivierung erfüllen, existieren bereits.

Die Finanzierung der digitalen Langzeitarchivierung von Forschungsdaten ist eine offene Frage, denn bislang gab es für Datenmanagement jenseits des Projektendes weder die notwendigen finanziellen Mittel, noch waren Antragsteller verpflichtet einen entsprechenden Plan vorzulegen. Hier tritt bei den Förderorganisationen inzwischen ein Umdenken ein. Durch die Umsetzung der „Empfehlung betreffend den Zugang zu Forschungsdaten aus öffentlicher Förderung“¹¹³ kann damit gerechnet werden, dass hier neue Möglichkeiten für den Aufbau von wissenschaftlichen Datenzentren und -archiven entstehen werden.

Technische Strategien

Voraussetzung für die digitale Langzeitarchivierung wissenschaftlicher Forschungsdaten ist, dass es vertrauenswürdige Archive gibt, die diese Aufgabe übernehmen können. Diese Aufgabe wird bereits in einigen Disziplinen von

108 Zerhouni (2006)

109 Bates et al. (2006); Sale (2006)

110 Severiens & Hilf (2006)

111 Lord & Macdonald (2003)

112 Lyon (2007)

113 OECD (2007)

Datenzentren übernommen und auch die Welt Datenzentren des International Council of Scientific Unions (ICSU WDCs) haben sich dieser Aufgabe verpflichtet. In den vielen Fällen, in denen es kein disziplinspezifisches Datenzentrum und –archiv gibt, fehlen Konzepte für eine digitale Langzeitarchivierung von wissenschaftlichen Forschungsdaten. Eine mögliche Lösung wäre, in Analogie zur Open Archive Initiative, für diese Daten lokale Institutional Repositories aufzubauen.¹¹⁴ Die Herausforderungen liegen dabei weniger bei den Archivsystemen, wo sie oft vermutet werden, sondern häufiger im Zusammenspiel der Prozesse des Managements von Forschungsdaten und der digitalen Langzeitarchivierung. So beziehen sich nur wenige Datenarchive in der Organisation ihrer Archivprozesse auf das OAIS-Referenzmodell (ISO 14721:2003)¹¹⁵, das die Prozesse der digitalen Langzeitarchivierung beschreibt.¹¹⁶

Besondere Herausforderungen an die digitale Langzeitarchivierung von Forschungsdaten erwachsen aus Grid- und eScience-Projekten, die sich auf den ersten Blick in vielen Aspekten nicht wesentlich von anderen Datenproduzierenden Forschungsprojekten unterscheiden. Die enorm großen Datenmengen, die in Grid-Projekten erzeugt und verarbeitet werden und die hohe Komplexität von Daten aus eScience-Projekten lassen jedoch vermuten, dass aus diesen Projekttypen neuartige Anforderungen an die digitale Langzeitarchivierung erwachsen.¹¹⁷ Gerade wegen dieser extremen Anforderungen an Prozessierungs- und Speicherressourcen und zusätzlichen Managementvorkehrungen durch Virtualisierung der Ressourcen sind Communities, die große Datenmengen erzeugen oder verarbeiten, in der Anwendung von Grid-Technologien vergleichsweise weit fortgeschritten. Astrophysik, Klimaforschung, biomedizinische Forschung, und andere Communities mit rechenintensiven Verfahren der Datenverarbeitung wenden bereits seit einiger Zeit Grid-Technologien an.

Die enorm großen Datenmengen erfordern von den Grid-Projekten konsistente Richtlinien für die Auswahl der Daten, die für lange Zeiträume archiviert werden sollen. Ähnlich wie in den Richtlinien des British Atmospheric Data Centre wird in den Projekten evaluiert, ob die Daten grundsätzlich und mit vertretbarem Aufwand neu generiert werden können, und ob die Daten in der vorliegenden Form nachnutzbar sind. Allerdings liegen die Herausforderungen an die Langzeitarchivierung von Forschungsdaten in Grid-Projekten weniger in den Datenmengen, sondern eher im neuartigen technologischen und organisa-

114 Lyon (2007)

115 OAIS (2002)

116 Lyon (2007)

117 Hey & Trefethen (2003)

torischen Rahmen (z.B. Virtuelle Organisationen, Authentifizierung und Autorisierung, semantische Bezüge, Metadaten).¹¹⁸

Langzeitarchive für Forschungsdaten und organisatorische Rahmenbedingungen in den Instituten und bei der Forschungsförderung sind notwendige Voraussetzungen für die digitale Langzeitarchivierung von wissenschaftlichen Forschungsdaten. Sie müssen aber auch durch technische Lösungen unterstützt werden, die die Mitwirkung durch die Wissenschaftler an der digitalen Langzeitarchivierung von wissenschaftlichen Forschungsdaten so einfach wie möglich gestalten, so dass sie sich möglichst nahtlos in die wissenschaftlichen Arbeitsabläufe einfügt. Ein Beispiel dafür ist die Beschreibung der Forschungsdaten durch Metadaten. Erstellen und Pflege von Metadaten stellt eine enorme Hürde dar, denn die notwendigen Metadatenprofile sind meist komplex, sie manuell zu erstellen ist aufwändig.¹¹⁹ In der Praxis hat sich gezeigt, dass das Management von Daten und Metadaten eine bessere Chance zum Erfolg hat, wenn das Erstellen und Pflegen von Metadaten weitgehend automatisiert ist. Ein hoher Grad an Technisierung des Datenmanagements erlaubt den Wissenschaftlern, sich ihrem eigentlichen Tätigkeitsschwerpunkt, der Forschung, zu widmen. In den vom Bundesministerium für Bildung und Forschung geförderten Projekten C3-Grid und TextGrid sind sowohl für die Naturwissenschaften, als auch für die Geisteswissenschaften vorbildliche Verfahren für die Erzeugung und Verwaltung von Metadaten entwickelt worden.¹²⁰

Während bereits die inhaltliche Beschreibung der zu archivierenden Daten durch Metadaten eine Hürde darstellt, kommen für eine spätere Nachnutzung weitere Probleme hinzu. Vielfach trifft man auf das Missverständnis, dass die Angabe des MIME-Type eine ausreichende Beschreibung des Dateiformats und seiner Nutzung sei. Ein Archivsystem müsste jedoch nicht nur den MIME-Type der archivierten Daten kennen, sondern auch deren semantische Struktur und ihr technisches Format. Die semantische Struktur maschinenlesbar zu dokumentieren ist eine Grundvoraussetzung für die in Zukunft geforderte Interoperabilität der Archivsysteme.¹²¹ Zusätzlich müssen sich die Archivbetreiber und ihre Nutzer darüber verständigen, welche Dateiformate archiviert werden, denn nicht jedes bei den Nutzern populäre Format ist für eine verlustfreie Langzeitarchivierung geeignet.¹²²

118 Klump (2008)

119 Robertson (2006)

120 Neuroth et al. (2007)

121 Klump (2008)

122 Lormant et al. (2005)

Ungeachtet des in der „Berliner Erklärung“ durch die Universitäten, Wissenschafts- und Forschungsförderungsorganisationen geleisteten Bekenntnisses zum offenen Zugang gibt es Gründe, warum manche Daten nicht offen zugänglich sein können. Aus diesem Grund sind Zugriffsbeschränkungen in der digitalen Langzeitarchivierung von Forschungsdaten ein wichtiges Thema. Die Zugriffsbeschränkungen dienen hierbei nicht primär der Sicherung von Verwertungsrechten, sondern sie sind entweder gesetzlich vorgeschrieben (Datenschutz) oder dienen dem Schutz von Personen oder Objekten, die durch eine Veröffentlichung der Daten Gefährdungen ausgesetzt würden. Für geschützte Datenobjekte müssen Verfahren und Policies entwickelt werden, die auch über lange Zeiträume hinweg zuverlässig die Zugriffsrechte regeln und schützen können.¹²³ Auch der Umgang mit „verwaisten“ Datenbeständen muss geregelt werden.

Zum Schutz der intellektuellen Leistung der Wissenschaftler sollten Daten in wissenschaftlichen Langzeitarchiven mit Lizenzen versehen sein, die die Bedingungen einer Nachnutzung regeln, ohne dadurch den wissenschaftlichen Erkenntnisgewinn zu behindern. Entsprechende Vorarbeiten sind bereits in den Projekten Creative Commons (CC) und Science Commons (SC) geleistet worden. Zusätzlich zur erwiesenen Praxistauglichkeit können die hier entwickelten Lizenzen auch maschinenlesbar hinterlegt werden, was eine künftige Nachnutzung deutlich vereinfacht. Die Diskussion, welche Lizenzen für Daten empfohlen werden sollten, ist noch offen.¹²⁴ Sie wird zusätzlich erschwert durch die rechtliche Auffassung im Urheberrecht, die die Schutzwürdigkeit von Daten stark einschränkt.

Nachnutzung von Daten

Keine der Infrastrukturen für eine digitale Langzeitarchivierung lässt sich dauerhaft betreiben, wenn es keine Nutzer gibt, denn erst wenn eine Nachfrage der Wissenschaft nach einer digitalen Langzeitarchivierung besteht, können dauerhafte Strukturen entstehen. Für die meisten Forschungsdaten gilt heute noch, dass die Nachfrage schon in den ersten Jahren stark abnimmt.¹²⁵ Dies gilt jedoch nicht für unwiederholbare Messungen wie z.B. Umweltmessdaten.¹²⁶ Im heutigen Wissenschaftsbetrieb sind der Gewinn an Distinktion und Reputation oft wichtige Motivationskräfte. Digitale Langzeitarchivierung muss als Praxis

123 Choi et al. (2006); Simmel (2004)

124 Uhler & Schröder (2007)

125 Severiens & Hilf (2006)

126 Pfeiffenberger (2007)

in der Wissenschaft verankert sein und im selbst verstandenen Eigeninteresse der Wissenschaftler liegen. Die wissenschaftliche Publikation ist dabei ein entscheidendes Medium.¹²⁷ Ein möglicher Anreiz, Daten zu veröffentlichen und dauerhaft zugänglich zu machen, ist es daher, die Veröffentlichung von Daten zu formalisieren und als Bestandteil des wissenschaftlichen Arbeitens zu institutionalisieren. Dazu ist nötig, dass die veröffentlichten Daten findbar, eindeutig referenzierbar und auf lange Zeit zugänglich sind. Allerdings werden Datenveröffentlichungen nur dann auch nachgenutzt und zitiert, wenn ihre Existenz den potenziellen Nutzern auch bekannt ist. Ein geeigneter Weg, Daten recherchierbar und zugänglich zu machen, ist ihre Integration in Fachportale und Bibliothekskataloge. Eine entscheidende Voraussetzung für die Zitierbarkeit von Daten ist, dass sie eindeutig und langfristig referenzierbar sind.¹²⁸

Da in der Praxis URLs nur kurzlebig sind, werden sie nicht als zuverlässige Referenzen angesehen. Persistente, global auflösbare Identifier, wie z.B. Digital Object Identifier (DOI) oder Universal Resource Names (URN) schließen diese Lücke.¹²⁹ In der bisherigen Praxis fehlten bisher wichtige Bestandteile, die eine nachhaltige Publikation von Daten möglich machen. Diese Defizite wurden im DFG-Projekt „Publikation und Zitierbarkeit wissenschaftlicher Primärdaten“ (STD-DOI) analysiert. Mit der Einführung von persistenten Identifikatoren für wissenschaftliche Datensätze wurden die Voraussetzungen für eine Publikation und Zitierbarkeit wissenschaftlicher Daten geschaffen.¹³⁰

Zusammenfassung

In der Einführung zum OAIS-Referenzmodell zur Langzeitarchivierung digitaler Objekte ist treffend formuliert worden, dass ein Archivsystem für digitale Objekte mehr ist als nur ein technisches System. Das OAIS-Referenzmodell beschreibt es als das Zusammenwirken von Menschen und Systemen mit dem Ziel der Langzeiterhaltung von digitalen Objekten für eine definierte Nutzergruppe.¹³¹ Die digitale Langzeitarchivierung von Forschungsdaten ist daher nicht allein eine technische Herausforderung, sondern muss in einen entsprechenden organisatorischen Rahmen eingebettet sein, der im Dialog mit der Wissenschaft gestaltet wird. Der wissenschaftliche Wert, Forschungsdaten für lange Zeit zu archivieren und zugänglich zu machen, ist erkannt worden. In dem Maße, wie

127 Pfeiffenberger & Klump (2006)

128 Klump et al. (2006)

129 Hilse & Kothe (2006); Klump et al. (2006)

130 Brase & Klump (2007)

131 OAIS (2002)

die Auswertung von Daten für die Forschung an Bedeutung zunimmt, wird sich auch der Umgang mit Daten in der Forschungspraxis und in der Langzeitarchivierung verändern.

Quellenangaben

- Barga, Roger, und Dennis B Gannon (2007): Scientific versus business workflows. In *Workflows for e-Science*, I. J. Taylor et al. (Hrsg.), S. 9-16, Springer-Verlag, London, Großbritannien.
- Bates, Melanie et al. (2006): *Rights and Rewards Project - Academic Survey Final Report*, JISC. Bath, Großbritannien.
- Berliner Erklärung (2003): Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>
- Brase, Jan, und Jens Klump (2007): Zitierfähige Datensätze: Primärdaten-Management durch DOIs, In *WisKom 2007 : Wissenschaftskommunikation der Zukunft ; 4. Konferenz der Zentralbibliothek, Forschungszentrum Jülich, 6. - 8. November 2007*, Bd. 18, R. Ball (Hrsg.), S. 159-167, Forschungszentrum Jülich, Jülich.
- Choi, Hee-Chul, et al. (2006): Trust Models for Community Aware Identity Management, in *WWW2006*, Edinburgh, Großbritannien.
- DFG (1998): *Sicherung guter wissenschaftlicher Praxis*, Deutsche Forschungsgemeinschaft, Bonn. http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf
- DFG (2009): *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*, Deutsche Forschungsgemeinschaft, Bonn. http://www.dfg.de/forschungsoerderung/wissenschaftliche_infrastruktur/lis/veroeffentlichungen/dokumentationen/download/ua_inf_empfehlungen_200901.pdf
- Hey, Tony, und Anne Trefethen (2003): e-Science and its implications, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 361(1809): 1809-1825, doi:10.1098/rsta.2003.1224.
- Hilse, Hans-Werner, und Jochen Kothe (2006): *Implementing Persistent Identifiers*, Consortium of European Research Libraries, London, Großbritannien.
- Klump, Jens (2008): *Anforderungen von e-Science und Grid-Technologie an die Archivierung wissenschaftlicher Daten*, Expertise, Kompetenznetzwerk Langzeitarchivierung (nestor), Frankfurt (Main). <http://nbn-resolving.de/urn:nbn:de:0008-2008040103>
- Klump, Jens, et al. (2006): Data publication in the Open Access Initiative, *Data Science Journal*, 5, 79-83, doi:doi:10.2481/dsj.5.79.

- Kluttig, Thekla (2008), Bericht über das DFG-Rundgespräch „Forschungsprimärdaten“ am 17.01.2008, Bonn, Germany. http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/forschungsprimaerdaten_0108.pdf
- Lord, Philipp, und Alison Macdonald (2003): *e-Science Curation Report - Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision*, JISC. Bath, Großbritannien.
- Lormant, Nicolas, et al. (2005): How to Evaluate the Ability of a File Format to Ensure Long-Term Preservation for Digital Information?, In: *Ensuring Long-term Preservation and Adding Value to Scientific and Technical data (PV 2005)*, S. 11, Edinburgh, Großbritannien. <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/003.pdf>
- Lyon, Liz (2007): *Dealing with Data: Roles, Rights, Responsibilities and Relationships*, consultancy report, UKOLN, Bath, Großbritannien. http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf
- Nature Redaktion (2006): A fair share, *Nature*, 444(7120), 653-654, doi:10.1038/444653b.
- Neuroth, Heike, et al. (2007): *Die D-Grid Initiative*, Universitätsverlag Göttingen, Göttingen. <http://resolver.sub.uni-goettingen.de/purl/?webdoc-1533>
- OAIS (2002): *Reference Model for an Open Archival Information System (OAIS). Blue Book*, Consultative Committee for Space Data Systems, Greenbelt, MD, USA.
- OECD (2004): *Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 January 2004 - Final Communiqué*, Communiqué, Organisation for Economic Co-operation and Development, Paris, Frankreich.
- OECD (2007): *OECD Principles and Guidelines for Access to Research Data from Public Funding*, Organisation for Economic Co-operation and Development, Paris, Frankreich.
- Pfeiffenberger, Hans (2007): Offener Zugang zu wissenschaftlichen Primärdaten, *Zeitschrift für Bibliothekswesen und Bibliographie*, 54(4-5), 207-210.
- Pfeiffenberger, Hans, und Jens Klump (2006): Offener Zugang zu Daten - Quantensprung in der Kooperation, *Wissenschaftsmanagement*, (Special 1), 12-13. http://oa.helmholtz.de/fileadmin/Links/Artikel/Wissenschafts_Management_Open_Access/Daten.pdf
- Robertson, R. John (2006): *Evaluation of metadata workflows for the Glasgow ePrints and DSpace services*, University of Strathclyde, Glasgow, Großbritannien.
- Sale, Arthur (2006): The acquisition of Open Access research articles, *First Monday*, 11(10).

- Severiens, Thomas, und Eberhard R Hilf (2006): *Langzeitarchivierung von Robdaten*, nestor - Kompetenznetzwerk Langzeitarchivierung, Frankfurt (Main). <http://nbn-resolving.de/urn:nbn:de:0008-20051114018>
- Simmel, Derek (2004): *TeraGrid Certificate Management and Authorization Policy*, policy, Pittsburgh Supercomputing Center, Carnegie Mellon University, University of Pittsburgh, Pittsburg, PA, USA. <http://www.teragrid.org/policy/TGCertPolicy-TG-5.pdf>
- Spittler, Gerd (1967), *Norm und Sanktion. Untersuchungen zum Sanktionsmechanismus*, Walter Verlag, Olten, Schweiz.
- Treloar, Andrew, et al. (2007): The Data Curation Continuum - Managing Data Objects in Institutional Repositories, *D-Lib Magazine*, 13(9/10), 13, doi:10.1045/september2007-treloar.
- Treloar, Andrew, und Cathrine Harboe-Ree (2008): Data management and the curation continuum: how the Monash experience is informing repository relationships. In: *VALA2008*, Melbourne, Australien. http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf
- Uhlir, Paul F, und Peter Schröder (2007): Open Data for Global Science, *Data Science Journal*, 6(Open Data Issue), OD36-53, doi:10.2481/dsj.6.OD36.
- Zerhouni, Elias A (2006): *Report on the NIH public access policy*, National Institute of Health, Bethesda, MD, USA. http://www.mlanet.org/government/gov_pdf/2006_nihrpt_pubaccessplcy.pdf