

Handbuch Forschungsdatenmanagement

Herausgegeben von
Stephan Büttner, Hans-Christoph Hobohm, Lars Müller

BOCK + HERCHEN Verlag
Bad Honnef
2011

Die Inhalte dieses Buches stehen auch als Online-Version zur Verfügung:
www.forschungsdatenmanagement.de

Die Onlineversion steht unter folgender Creative-Common-Lizenz:

„Attribution-NonCommercial-ShareAlike 3.0 Unported“

<http://creativecommons.org/licenses/by-nc-sa/3.0/>



ISBN 978-3-88347-283-6

BOCK+HERCHEN Verlag, Bad Honnef

Printed in Germany

3.4 Archivierung von Forschungsdaten

Erich Weichselgartner, Armin Günther, Ina Dehnhard

Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)

Forschungsprojekte enden in der Regel nicht, weil das Forschungsziel bereits komplett erreicht, die erhobenen Daten umfassend ausgewertet und die Befunde ausführlich kommuniziert worden sind. Vielmehr wird die Arbeit eingestellt, weil die Finanzierungen ausgelaufen, das (befristet beschäftigte) Personal abgewandert oder andere Forschungsthemen in den Vordergrund getreten sind. Selbst bei planmäßigem Ende empirischer Forschungsprojekte können die erhobenen Daten in der Regel nicht erschöpfend ausgewertet werden. Gründe hierfür sind begrenzte Ressourcen und der zielorientierte Charakter von Forschungsvorhaben. Datenanalysen außerhalb des engeren Zielfokus und abweichende Fragestellungen, evtl. sogar aus anderen Disziplinen, können im Sinne dieser Forschungsökonomie kaum Beachtung finden. Zudem bleibt solchen Re- oder Metaanalysen ein Riegel vorgeschoben, wenn die Daten von Forschern für andere unzugänglich aufbewahrt werden, wie es in den meisten Disziplinen nach wie vor gängige Praxis ist (vgl. PARSE.Insight, 2009, S. 33–34, mit Ergebnissen einer europaweiten Umfrage). Bei teuren oder einmaligen Datenerhebungen ist dies ein Ärgernis zum Schaden der gesamten Wissenschaftsgemeinschaft sowie der Gesellschaft als Ganzes. Dabei gibt es althergebrachte Gegenwürfe. Zum Beispiel hat die Reanalyse astronomischer und geodätischer Forschungsdaten aus dem Jahre 1796 zur Entdeckung psychologisch-kognitions-wissenschaftlicher Gesetzmäßigkeiten geführt, die wiederum für die Ergonomie große Bedeutung haben: Die persönliche Gleichung (Bessel, 1822) und das *Law of Prior Entry* (Titchener, 1908).

Während im 19. Jh. Forschungsdaten noch im Anhang von Publikationen abgedruckt wurden (z. B. Donders, 1868), gab man diese Praxis im 20. Jh. auf, nicht zuletzt aus Gründen der Praktikabilität bei großen Datenmengen. Alternativ haben sich Forscher verpflichtet, die Daten für eine bestimmte Frist aufzubewahren (die Deutsche Forschungsgemeinschaft empfiehlt 10 Jahre; DFG, 1998) und bei Bedarf interessierten Fachkolleginnen zugänglich zu machen (in der Psychologie z. B. gemäß den ethischen Richtlinien der Deutschen Gesellschaft für Psychologie (DGPs); DGPs, 2004). In der Praxis ist es mit diesen Grundsätzen aber nicht weit her, wie Wicherts, Borsboom, Kats und Molenaar (2006) feststellen mussten, die nur in jedem fünften Fall der Datennachfrage erfolgreich waren. Dieser Missstand hat zu einem kritischen Editorial im Wissenschaftsmagazin *Nature* (Nature, 2006) geführt und mit dazu beigetragen, über die Fachgrenzen hinweg Aufmerksamkeit auf die Archivierung und Zugänglichkeit von Forschungsdaten zu lenken. Dazu trug auch die digitale Revolution bei, durch die sich die Rahmenbedingungen änderten und die nachhaltige Sicherung und

Bereitstellung von Forschungsdaten weltweit zu einem wichtigen Thema für die wissenschaftliche Infrastruktur wurde.

Im Folgenden soll aufgezeigt werden, warum die Archivierung und Bereitstellung von Forschungsdaten wichtig ist, wo Probleme liegen, welche Selektionskriterien angemessen sind, welche Herausforderungen zu bewältigen sind und welche Trends sich abzeichnen.

3.4.1 Nutzen von Datenarchivierung und -bereitstellung

Intrinsische Gründe. Datenerhebungen sind mitunter sehr teuer und der geleistete Einsatz soll einen möglichst hohen wissenschaftlichen Gewinn bringen, d.h. die Daten sollen möglichst *umfassend ausgewertet* werden. Ein weiterer Grund, der gerade bei Untersuchungen an Mensch und Tier relevant wird, ist die *Vermeidung redundanter Datenerhebungen*, da diese immer auch eine Belastung sein können (vgl. Ethical Principles of Psychologists and Code of Conduct; APA, 1992). Darüber hinaus sind manche Daten einmalig und können *nicht repliziert* werden (z. B. sozial- und kulturwissenschaftliche Daten, die unmittelbar nach Beendigung der deutschen Teilung erhoben wurden). Weitere Gründe sind die Möglichkeiten der Analyse von Datensätzen unter *neuen Fragestellungen* und Perspektiven, der Reanalyse von Datensätzen mit *anderen Analysemethoden* (u.U. solche, die zum Zeitpunkt der Datenerhebung noch nicht bekannt waren), der *vergleichenden Analyse* verschiedener Datensätze zur Ermittlung der Robustheit von Ergebnissen oder zur Ermittlung der Generalisierbarkeit empirischer Gesetze, der *Ermittlung historischer Veränderungen* und der präziseren *Replikation von Untersuchungen*, die basierend auf den publizierten Forschungsberichten nur in Annäherung möglich wären.

Extrinsische Gründe. Für eine Datenarchivierung sprechen zudem die *Erfüllung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis* und die Möglichkeit der *Überprüfung* publizierter Datenanalysen. Für Letzteres besteht gerade im Zusammenhang mit der Aufdeckung von Fällen wissenschaftlichen Fehlverhaltens durchaus ein Bedarf. Des Weiteren soll die Bereitstellung von Forschungsdaten der *Verstärkung des wissenschaftlichen Austauschs* dienen und die Rezeption der Forschung durch die Fachöffentlichkeit unterstützen. Nach einer Untersuchung von Piwowar und Chapman (2008) haben Zeitschriften mit einer expliziten Richtlinie zur Datenbereitstellung zudem einen höheren *Impact Factor*. Nicht zuletzt soll das wissenschaftliche Gratifikationssystem die Arbeit der Datenerhebung stärker (gerechter) würdigen und Forschungsdaten sollen deshalb *zitierbar* werden (siehe das DataCite-Projekt; DataCite, 2011).

3.4.2 Unterschiede zwischen Disziplinen

Grundsätzlich besteht über die Wissenschaftsdisziplinen hinweg weitgehend Einigkeit, dass Forschungsdaten erhalten und anderen zugänglich gemacht werden sollen. In der Praxis des Forschungsdatenmanagements aber, so haben vergleichende Untersuchungen etwa des *Research Information Network* (RIN, 2008) oder des *Digital Curation Centre* (Key Perspectives, 2010) gezeigt, bestehen große Unterschiede zwischen den Disziplinen oder innerhalb einer Disziplin zwischen unterschiedlichen Forschungsfeldern.

In der Astronomie existiert nicht nur eine hohe Bereitschaft der Forschenden, ihre Daten zu publizieren, sondern es besteht auch eine gut entwickelte Infrastruktur (*Open Access* Datenbanken bzw. Datenzentren). Ähnliches gilt für die Genforschung. Demgegenüber stellt sich die Situation in den Wirtschafts-, Sozial- und Verhaltenswissenschaften deutlich anders dar. Hier ist die Archivierung und Veröffentlichung von Forschungsdaten eher die Ausnahme, denn die Regel. Zwar gibt es auch in diesen Disziplinen große, häufig nationale Umfragestudien, bei denen die Daten gut dokumentiert, veröffentlicht und intensiv von anderen Forschenden genutzt werden (wie z. B. das Sozio-oekonomische Panel (SOEP, o. J.). Dem steht aber eine sehr viel größere Zahl von kleineren Projekten und Studien gegenüber, bei denen die Forschungsdaten lediglich als Basis für Publikationen der projektbeteiligten Forscher dienen, darüber hinaus aber nicht systematisch archiviert und öffentlich zugänglich gemacht werden.

3.4.3 Anreizstruktur

Wie lassen sich diese Unterschiede in der Praxis der Datenarchivierung und -veröffentlichung erklären und – wenn man die Gründe hierfür kennt – verringern? Von grundlegender Bedeutung dürfte die Anreizstruktur der jeweiligen Disziplin oder des jeweiligen Forschungsfeldes sein, d. h. die zu erwartenden „Kosten“ und „Nutzen“, die mit der Archivierung und Veröffentlichung von Forschungsdaten (in der Erwartung der Forschenden) verbunden sind. Zu den *Kosten* gehören:

- Zeit- und Geld-Aufwand für die Archivierung und Veröffentlichung von Forschungsdaten
- Nachteile im wissenschaftlichen Wettbewerb, indem andere die Forschungsdaten wissenschaftlich auswerten (vgl. Haeussler, 2010)
- reputationsschädigende Aufdeckung methodischer Unzulänglichkeiten der Datenauswertung
- Nachteile bei der ökonomischen Nutzung von Forschungsdaten
- rechtliche Probleme durch die Veröffentlichung von Daten (Verletzung des Datenschutzes oder von Eigentumsrechten z. B. bei geschützten Messverfahren)

- die Beschäftigung mit Dingen (wie der „Verwaltung“ von Daten), die dem Selbstverständnis der Forschenden nicht entsprechen

Den Kosten stehen *positive Anreize*, die eigenen Forschungsdaten zu archivieren und zu veröffentlichen, gegenüber. Diese müssen nicht einmal altruistischer Natur sein (Dienst am wissenschaftlichen Fortschritt), sondern können durchaus auch im Eigeninteresse der Forschenden liegen:

- Zugang zu Ressourcen (z. B. Fördermitteln)
- direkter wissenschaftlicher Reputationsgewinn durch die Veröffentlichung von qualitativ hochwertigen Forschungsdaten
- erhöhte Sichtbarkeit der eigenen (weiteren) Forschungsarbeit
- Anerkennung in der *scientific community* (Erfüllung sozialer Normen)
- Eröffnen neuer Kooperations- und Publikationsmöglichkeiten (z. B. Co-Autorenschaft mit Datennutzern)
- wissenschaftliche Anregungen durch die gemeinschaftliche Arbeit mit denselben Daten

Wie die Bilanz von Kosten und Nutzen ausfällt, hängt wesentlich von den institutionellen Strukturen einer Disziplin oder eines Forschungsfeldes ab, wobei auch regionale Unterschiede zum Tragen kommen. In dem Maße, in dem beispielsweise der Zugang zu Fördermitteln von der verpflichtenden Archivierung der in einem Förderprojekt erhobenen Forschungsdaten abhängig gemacht wird, verschiebt sich die Relation von den Kosten (Zeit- und Geldaufwand der Archivierung) hin zum Nutzen (Zugang zu Ressourcen/Fördermitteln). Beispielsweise verlangt das amerikanische nationale Gesundheitsinstitut (NIH) ab Fördersummen von 0,5 Mio. Dollar einen Plan zur Datenarchivierung und -weitergabe (NIH, 2003). Ebenso verschiebt sich die Kosten-Nutzen Relation in dem Maße, in dem Infrastruktureinrichtungen und Dienstleistungen zur Verfügung stehen, die Forschende bei der Archivierung von Daten unterstützen und damit Arbeitskosten verringern. Hiermit einher geht die Entwicklung von Normen und Standards, von Prozessen und *best practices*, von Werkzeugen und Maschinen zur Datenarchivierung, so dass Forschende, die „willig“ sind, auch auf Strukturen zurückgreifen können, die ihnen überhaupt erst eine sinnvolle Datenarchivierung ermöglichen.

Allerdings ist die Existenz einer solchen Infrastruktur alleine noch keine Garantie dafür, dass sie auch genutzt wird, wie RIN (2008, S. 14) für den Fall der *Social and Public Health Sciences* feststellt. Auch „weichere“ Faktoren spielen eine nicht unwesentliche Rolle. In vielen Disziplinen gibt es für die Veröffentlichung von Daten zwar keine in formalisierten Evaluierungsverfahren institutionalisierte Anerkennung als wissenschaftliche Leistung, die der Veröffentlichung eines Aufsatzes vergleichbar wäre. Aber in einigen Disziplinen, wie der Astronomie oder der Genforschung, gibt es offenbar eine klare Erwartung,

dass Forschungsdaten dennoch archiviert und veröffentlicht werden. Wer diese Erwartung nicht erfüllt, gefährdet sein Ansehen in der *community*. In anderen Disziplinen wie z. B. der Psychologie, gibt es einen solchen sozialen Druck nicht. Hier ist es üblich und normal, seine Daten nicht zu veröffentlichen.

Auch die Art der Daten, die in einer Disziplin oder einem Forschungsfeld vorwiegend produziert werden, spielt eine wichtige Rolle. Dort, wo die Deskription natürlicher oder sozialer Systeme und Prozesse im Vordergrund steht, sind die entsprechenden Forschungsdaten oftmals für die *community* eine unverzichtbare Arbeitsgrundlage und werden entsprechend auch nachgefragt. Dort, wo eher das Testen von Hypothesen oder Modellen im Vordergrund steht, wird die Archivierung und Veröffentlichung von Forschungsdaten vielfach als weniger wichtig angesehen. In der Klimaforschung sind beide Fälle anzutreffen: Während bei Beobachtungsdaten aus der Fernerkundung zumindest die Erwartung besteht, dass die Daten der *community* dauerhaft zur Verfügung gestellt werden, wird für Daten aus Modellrechnungen eine Archivierung nur begrenzt (auch in zeitlicher Hinsicht!) für sinnvoll angesehen (vgl. RIN, 2008, S. 64).

3.4.4 Selektion

Kaum eine Archiv- oder Forschungseinrichtung sieht sich in der Lage, alle in der jeweiligen Disziplin produzierten Daten aufzubewahren, so dass sich die Frage nach der Selektion stellt. Besonders in Wissenschaften wie der Astronomie, den Geowissenschaften oder der Hochenergiephysik, bei denen in machen Forschungsstationen oder Datenzentren täglich mehrere Terabyte Daten eingehen, macht der große Datenumfang Selektionskriterien erforderlich. Abgesehen von der Datenmenge liegen Beschränkungen für die Archivierung von Daten im Aufbereitungsaufwand begründet. In den meisten Wissenschaften ist eine ausführliche Beschreibung der zu archivierenden Daten unverzichtbar, damit die Daten für eine spätere Nachnutzung interpretierbar bleiben. Diese Datendokumentationen erfordern hohe personelle und z. T. auch technische Ressourcen, wenn die Apparaturen und Messgeräte der Datenerhebung ebenfalls archiviert werden müssen.

Leitendes Auswahlkriterium sollte die Bedeutsamkeit der Daten für potentielle Nachnutzer sein. Dementsprechend enthalten auch die Selektionsstrategien verschiedener Datenzentren und -archive immer den Aspekt des Nutzens für nachfolgende Analysen. Das *UK Data Archive* (2009, S. 2f.) stellt als grundlegendes Auswahlprinzip fest: „The UKDA collects data, information and other electronic resources of long-term interest and use across the range of social science and historical disciplines.“ Als Hinweis auf ein besonderes Analysepotential wird ein großer Stichprobenumfang genannt. Das *Earth Resources Observation and Science Center* (EROS, 2007) stellt als zusätzliches Prüfkriterium die Frage, welche Konsequenzen es hätte, wenn gerade diese Daten nicht archiviert

werden würden. Studien unter historisch einmaligen Rahmenbedingungen, wie z. B. Untersuchungen im Zusammenhang zur deutschen Einheit oder des Ausbruchs des Vulkans *Eyjafjallajökull*, sind aufgrund ihrer Einmaligkeit prädestiniert für die Archivierung. PsychData (o.J.), ein Forschungsdatenzentrum für die Psychologie, definiert ebenfalls Kriterien in Bezug auf die Bedeutsamkeit von Daten im Forschungsfeld der Psychologie: So wird psychologischen Längsschnittstudien und umfangreichen Querschnittstudien aufgrund ihres Datenumfanges und ihrer Repräsentativität ein besonders hohes Analysepotential zugesprochen (vgl. Weichselgartner, 2008). Auch wenn die Bedeutsamkeit von Daten ein unstrittiges Auswahlkriterium zu sein scheint, so bereitet seine Einschätzung doch Schwierigkeit, insbesondere wenn man den zukünftigen wissenschaftlichen Wert von Forschungsdaten antizipieren möchte. Wie kann heute beurteilt werden, was morgen (in einigen Jahren, Jahrzehnten oder sogar darüber hinaus) prominente Forschungsthemen sein werden? Deswegen verwenden Datenarchive zusätzliche, konkreter definierte Mechanismen für die Datenauswahl. Da die meisten Archivangebote disziplinspezifisch und teilweise innerhalb der Disziplinen noch themenspezifisch ausgerichtet sind, ergibt sich eine gewisse Vorauswahl der Daten bereits hier. Die Erweiterung und Passung eines neuen Datensatzes zu bereits vorhandenen Datenkollektionen ist ebenso von Belang wie die Abdeckung eines Themengebiets, zu dem bisher noch keine archivierten Daten vorhanden sind (siehe UK Data Archive, 2009, S. 3 f. und EROS, 2007). Weitere Selektionskriterien für die Aufnahme von Datensätzen können geografische (z. B. Deutschland, Europa) oder zeitlich-historische (z. B. Periode Weimarer Republik) Beschränkungen bei der Auswahl von archivierungswürdigen Datensätzen sein.

Wesentliches Auswahlkriterium stellt aber disziplinübergreifend die Qualität und Beschaffenheit der zu archivierenden Daten dar. Zum einen wird anhand der Art und der Formate der Daten entschieden, ob entsprechende Techniken und *Know-How* für ihre Aufbereitung und Archivierung vorhanden sind. Zum anderen stellt die Dokumentation der Daten eine wesentliche Voraussetzung für ihre Aufbereitung und Bereitstellung dar. Eine unzureichende und lückenhafte Datendokumentation bedeutet immer, dass Daten für die Archivierung abgelehnt werden, da die Daten nicht mehr interpretierbar und analysierbar sind. Weitere Qualitätskriterien stellen die Authentizität, Integrität und Konsistenz der Daten dar, welche vor Übernahme in ein Archiv überprüft werden.

Ebenfalls zu berücksichtigen bei der Auswahl von Datensätzen sind rechtliche Aspekte. Fragen des Datenschutzes spielen besonders in den Sozial- und Wirtschaftswissenschaften eine wesentliche Rolle bei der Datenselektion (vgl. APA, 1992). Durch Mechanismen wie Anonymisierungsverfahren oder beschränktem Datenzugriff kann in vielen Fällen eine Archivierung und Bereitstellung zwar realisiert werden, jedoch teilweise nicht ohne Informationsverluste. Mögliche

rechtliche Ansprüche auf die Daten müssen eindeutig geklärt sein, bevor die Übergabe von Daten an Datenzentren möglich ist.

3.4.5 Herausforderungen

Datendokumentation. Bei der Langzeiterhaltung von Forschungsdaten stellt eine auf langfristige Interpretierbarkeit ausgerichtete Dokumentation der Daten eine wesentliche Notwendigkeit dar. Forschungsdaten selbst sind ohne eine zugehörige Dokumentation, welche die Erhebungsbedingungen, den Kontext und die Forschungsinstrumente beschreibt, nicht interpretierbar. Je nach Forschungskontext und untersuchtem Objekt sind die Anforderungen an den Umfang und die Detailliertheit der Dokumentation unterschiedlich. Besonders in solchen Disziplinen, in denen es sich um einmalige, nicht wiederholbare Ereignisse handelt, die mit speziellen Messverfahren erfasst werden, ist eine genaue Beschreibung des Forschungskontextes notwendig. So muss in der Ökologie für eine adäquate Interpretation der Daten vielfach der Erhebungskontext detailliert beschrieben werden (vgl. Zimmerman, 2008), während in der Chemie eine stärker standardisierte Beschreibung möglich ist.

In vielen Disziplinen ist die Datendokumentation nicht in den Forschungs- und Publikationsablauf integriert. Datendokumentationen bedeuten einen hohen zusätzlichen Zeit- und Arbeitsaufwand, der meist nicht entsprechend „entlohnt“ wird. Sowohl fehlende wissenschaftliche Honorierungen (*credits, incentives*) als auch die unzureichende Einplanung der Datendokumentation in der Budgetplanung eines Forschungsvorhabens führen zu einer Vernachlässigung des Datenmanagements im Forschungsalltag. Der Beginn neuer Forschungsprojekte erscheint attraktiver als eine Aufbereitung alter Forschungsprojekte für die Archivierung und Nachnutzung (Esanu et al., 2004). Hinzu kommen in einigen Disziplinen das Fehlen einheitlicher allgemein anerkannter Dokumentationsstandards sowie die fehlende Vermittlung von Standards und Praktiken des Forschungsdatenmanagements in der wissenschaftlichen Ausbildung. Die Implementierung des Dokumentationsprozesses in den eigentlichen Forschungsprozess stellt also eine besondere Herausforderung dar. Als optimal hat sich eine forschungsbegleitende Dokumentation von Beginn des Forschungsvorhabens an herausgestellt, da alle notwendigen Informationen noch direkt verfügbar sind. *Tools*, die die Datendokumentation bereits während des Forschungsprozesses unterstützen, kommt hier eine wichtige Funktion zu (vgl. z. B. Dehnhard und Weiland (2011) für das Dokumentationstool des Datenzentrums PsychData).

Entwicklung von Standards. Doch auch eine für sich genommen „gute“ Datendokumentation, die die nach gegenwärtigem Wissenstand wesentlichen Metadaten zum Verständnis der Forschungsdaten enthält, ist möglicherweise nur von begrenztem Nutzen, wenn diese Dokumentation nicht in einer standardisierten Form erfolgt. Bei dem rapide wachsenden Volumen an archivierten

und veröffentlichten Forschungsdaten wird eine Standardisierung der Datendokumentation zunehmend wichtiger, damit diese Daten zunächst archiviert und dann gefunden, mit Daten aus anderen Forschungsprojekten verknüpft und wiederverwendet werden können. Es sind daher in verschiedenen Disziplinen Initiativen entstanden mit dem Ziel, für bestimmte Forschungsgebiete Datenstandards zu entwickeln, die festlegen, welche Informationen mindestens bei einer Datendokumentation in diesem Forschungsbereich enthalten sein sollen („*minimum reporting standards*“ oder „*minimum information specification*“) und wie diese Informationen dokumentiert werden sollen (vgl. z. B. für den Bereich der Biologie OMICS, 2006).

Entwicklung von Werkzeugen. Standardisierung ist ein Aspekt einer Herausforderung, die ebenfalls mit dem wachsenden Volumen und der wachsenden Komplexität von Daten verbunden ist: die Entwicklung von *Tools* und Verfahren, genau die Daten zu finden, die in einem Forschungskontext gebraucht werden können. Bereits bei einer einzelnen psychologischen oder sozialwissenschaftlichen Studie beispielsweise können Daten in erheblichem Umfang und Komplexität (Anzahl und Struktur der erhobenen Variablen, abgeleitete Daten usw.) anfallen, die es einem projektfremden Nutzer schwer machen, diese Daten zu entdecken, abzurufen und zu verwenden. Riesige Mengen an Daten anzuhäufen macht nur dann Sinn, wenn es Hilfsmittel gibt, die in diesen Datenmengen enthaltenen Schätze auch zu heben – ansonsten entstehen lediglich gigantische Datenfriedhöfe.

3.4.6 Ausblick

Ein wichtiger Trend, zumindest in der Theorie des Datenmanagements, ist die Entwicklung weg von einer am Ende des Forschungsprozesses stehenden oder gar nachträglichen Archivierung von Forschungsdaten hin zu einem in den Forschungsprozess integrierten *data-lifecycle* Management. Die Vision hierbei ist, dass bereits während der Konzeptions- und Durchführungsphase eines Forschungsprojektes die Dokumentation, Archivierung und Veröffentlichung von Forschungsdaten vorbereitet wird, indem beispielsweise die für die Dokumentation erforderlichen Metadaten (manuell, semi- oder vollautomatisch) erfasst und standardkonform aufbereitet werden. Eine eigene Archivierungsphase kann sich dann weitgehend auf die Übergabe der Daten und Datendokumentation an ein Forschungsdatenzentrum beschränken.

Es hat sich immer wieder gezeigt, dass die Auseinandersetzung mit den Erfordernissen der Datenarchivierung und -veröffentlichung nicht erst am Ende eines Forschungsprozesses erfolgen sollte. Nicht selten lassen sich wichtige Aspekte der Datenerhebung nicht mehr rekonstruieren, teilnehmende Forscher sind nicht mehr greifbar oder es wurde auch nur vergessen, sich die für die Veröffentli-

chung der Daten erforderlichen Rechte zu besorgen (z. B. bei Untersuchungen mit Menschen das Einverständnis zur Veröffentlichung der Forschungsdaten).

Die sich abzeichnende zunehmende Verfügbarkeit von Forschungsdaten wird Auswirkungen auf das wissenschaftliche Arbeiten quer über alle Disziplinen haben. Rechenbetonte Methoden (*computational methods*) könnten gleichberechtigt neben die klassische (Hypothesen geleitete) Forschung treten, denn mit Hilfe der enormen verfügbaren Datenmengen können beliebig viele Hypothesen „durchprobiert“ werden. Der Begriff „*Data-Driven Science*“ wurde als viertes wissenschaftliches Paradigma von Jim Gray 2007 (Hey, Tansley & Tolle 2009) ins Spiel gebracht (die anderen drei sind „Empirie“, „Theorie“ und „Simulation“). Die großen Herausforderungen des 21. Jh. wie Energie- und Umweltproblematiken sind möglicherweise nur auf diese Weise bewältigbar.

Literaturhinweise

- APA (American Psychological Association), 1992. Ethical principles of psychologists and code of conduct. *American Psychologist*, 47(12), S. 1597–1611.
- Bessel, F. W., 1823. *Astronomische Beobachtungen auf der Königlichen Universitäts-Sternwarte in Königsberg. 8. Abtheilung vom 1. Januar bis 31. December 1822*. Königsberg, III–VIII.
- DataCite, 2011. *Welcome to DataCite*. Online: <http://datacite.org> [Zugriff am 17.02.2011].
- Dehnhard, I. & Weiland, P., 2011. Toolbasierte Datendokumentation in der Psychologie. In: J. Griesbaum, T. Mandl & C. Womser-Hacker, Hrsg. *Information und Wissen: global, sozial und frei? Proceedings des 12. Internationalen Symposiums für Informationswissenschaft*. Boizenburg: Hülsbusch, S. 74–84.
- DGPs (Deutsche Gesellschaft für Psychologie), 2004. *Revision der auf die Forschung bezogenen ethischen Richtlinien*. Online: <http://www.dgps.de/dgps/aufgaben/ethikrl2004.pdf> [Zugriff am 17.02.2011].
- DFG (Deutsche Forschungsgemeinschaft), 1998. *Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“*. Weinheim: Wiley-VCH.
- Donders, F. C., 1868. Die Schnelligkeit psychischer Prozesse. *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin*. Berlin, S. 657–681.
- EROS (Earth Resources Observation and Science), 2007. *Records appraisal tool*. Online: <http://eros.usgs.gov/government/ratool/> [Zugriff am 17.02.2011].
- Esanua, J. Davidson, J. Ross, S. & Anderson, W., 2004. Selection, appraisal, and retention of digital scientific data: Highlights of an EPRANET/CODATA Workshop. *Data Science Journal*. Online: http://www.jstage.jst.go.jp/article/dsj/3/0/227/_pdf [Zugriff am 17.02.2011].
- Haeussler, C., 2010. *Information-sharing in academia and the industry: A comparative study*. (RatSWD Working Paper No. 154) Online: http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_154.pdf [Zugriff am 17.02.2011].
- Hey, T. Tansley, T. & Tolle, K., 2009. Jim Gray on eScience: A Transformed Scientific Method. Based on the transcript of a talk given by Jim Gray to the NRC-CSTB in Mountain View, CA, on January 11, 2007. In: A. Hey, St.

- Tansley & K.M. Tolle, 2009. *The Fourth Paradigm Data-Intensive Scientific Discovery*. Redmond, Wash.: Microsoft Research, S. xvii-xxxii
- Key Perspectives, 2010. *Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study*. Edinburgh: Digital Curation Centre. Online: <http://www.dcc.ac.uk/scarp> [Zugriff am 17.02.2011].
- Nature, 2006. A fair share. *Nature*, 444 (7 December 2006), S. 653–654.
- NIH (National Institutes of Health), 2003. *NIH data sharing policy and implementation guidance*. (Stand: 5.3.2003) Online: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm [Zugriff am 17.02.2011].
- OMICS, 2006. *OMICS A Journal of Integrative Biology*, Vol. 10 (Special Issue on Data Standards).
- PARSE.Insight, 2009. *Survey report*. (D3.4) Online: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf [Zugriff am 17.02.2011].
- Piwowar, H.A. & Chapman, W.W., 2008. A review of journal policies for sharing research data. In: *Proceedings of the 12th International Conference on Electronic Publishing*. Toronto, Kanada 25.-27. Juni 2008. Online: http://elpub.scix.net/data/works/att/001_elpub2008.content.pdf [Zugriff am 17.02.2011].
- PsychData, o.J. *PsychData – Startseite*. Forschungsdaten für die Psychologie. Online: <http://www.psychdata.de/> [Zugriff am 17.02.2011].
- RIN (Research Information Network), 2008. *To share or not to share: Publication and quality assurance of research data outputs. Annex: detailed findings for the eight research areas*. Research Information Network. Online: <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-annex.pdf> [Zugriff am 17.02.2011].
- SOEP (Sozio-oekonomisches Panel), o.J. DIW Berlin. *Startseite SOEP*. Online: <http://www.diw.de/soep> [Zugriff am 17.02.2011].
- Titchener, E. B., 1908. *Lectures on the elementary psychology of feeling and attention*. New York: Macmillan.
- UK Data Archive, 2009. *UK Data Archive collections development policy*. Online: <http://www.esds.ac.uk/news/publications/UKDACollectionsDevPolicy.pdf> [Zugriff am 17.02.2011].

- Weichselgartner, E., 2008. Fünf Jahre Primärdatenarchivierung in der Psychologie: Ein Erfahrungsbericht. DGI (Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis), 30. *DGI Online-Tagung 2008*. Frankfurt am Main, Deutschland 15.-17. Okt. 2008. Online: <http://www.weichselgartner.de/edu/downloads/dgi2008.pdf> [Zugriff am 17.02.2011].
- Wicherts, J. M. Borsboom, D. Kats, J. & Molenaar, D., 2006. The poor availability of psychological research data for reanalysis. *American Psychologist*, (61), S. 726–728.
- Zimmerman, A. S., 2008. New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science Technology Human Values*, (33), S. 631-652.