

# Toolbasierte Datendokumentation in der Psychologie

*Ina Dehnhard, Peter Weiland*

Leibniz-Zentrum für Psychologische Information und Dokumentation  
D-54286 Trier  
E-Mail: dehnhard@zpid, weiland@zpid.de

## **Zusammenfassung**

Das Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) stellt mit PsychData seit 2002 ein auf Forschende in der Psychologie ausgerichtetes Datenarchiv zur Verfügung. In diesem Beitrag werden die Herausforderungen der Forschungsdatendokumentation im Fach Psychologie dargestellt, und es wird ein im Rahmen des Projekts entwickeltes Werkzeug vorgestellt, das insbesondere in der forschungsbegleitenden Dokumentation eingesetzt werden kann.

## **Abstract**

Since 2002, the Leibniz Institute for Psychology Information is operating a data archive supporting researchers in psychology. This paper discusses the challenges of documenting research data in psychology and related disciplines. Moreover, a Web application is presented which has been developed for data documentation during the research process.

## **1 Einleitung**

Psychologie versteht sich als empirische Wissenschaft, deren Erkenntnisfortschritt auf der Gewinnung und Verarbeitung von Forschungsdaten beruht (Erdfelder, 1994). Die aufgrund der Daten ermittelten Ergebnisse werden dann in Publikationen der Forschungsgemeinschaft zur Verfügung gestellt. Während früher im Anhang einer Publikation noch die Daten veröffentlicht wurden (z.B. Donders, 1868) und so also direkt verfügbar waren, ist es heute

schwieriger, an die Datenbasis einer Veröffentlichung zu gelangen (Wicherts et al. 2006). Trotz des enormen Stellenwerts, der der empirischen Grundlage psychologischer Forschung beigemessen wird, erhalten die Forschungsdaten weniger Beachtung. Dies wird auch an der Tatsache deutlich, dass bisher die Veröffentlichung von Daten Wissenschaftlern keine Reputation bringt, sondern nur darauf aufbauende Publikationen (vgl. Wagner, 2010, Klump et al., 2006).

Allerdings lässt sich in Forschungskultur und -politik ein Umdenken feststellen. Die Vorteile des Data Sharing, wie z.B. ökonomischer Nutzen, Qualitätssicherung, Potential für Reanalysen unter neuen Fragestellungen oder mit anderen statistischen Methoden, werden zunehmend erkannt (vgl. Bengel & Wittman, 1982, Wicherts et al. 2006) und führen zu ersten forschungspolitischen und infrastrukturellen Maßnahmen. So gibt die Deutsche Forschungsgemeinschaft (DFG) in ihrer Denkschrift *Vorschläge zur Sicherung guter wissenschaftlicher Praxis* die Empfehlung, dass „Primärdaten als Grundlagen für Veröffentlichungen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden sollen“ (DFG, 1998).

Über die reine Aufbewahrungspflicht hinaus, sollten Daten auch für andere Wissenschaftler auffindbar und zugänglich sein. Die Verbesserung und Erweiterung der Dateninfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften stellt daher das erklärte Arbeitsziel des Rates für Sozial- und Wirtschaftsdaten (RatSWD)<sup>1</sup> dar. Der RatSWD ist ein unabhängiges Gremium, das sich aus empirisch arbeitenden Wissenschaftlern von Universitäten, Hochschulen und anderen Forschungseinrichtungen unabhängiger wissenschaftlicher Forschung sowie von Vertreterinnen und Vertretern wichtiger Datenproduzenten zusammensetzt. Seine Aufgaben liegen in der Optimierung des Datenzugangs und der Datennutzung sowie in der Interessenvermittlung zwischen Wissenschaft und Datenproduzenten.

Speziell für das Fach Psychologie bietet das Datenarchiv PsychData<sup>2</sup> ein Serviceangebot zur Unterstützung von Forschenden im Primär- und

---

<sup>1</sup> <http://www.ratswd.de/>, letzter Zugriff am 25.10.2010.

<sup>2</sup> <http://psychdata.de/>, letzter Zugriff am 04.11.2010

Sekundärforschungskontext an (vgl. Ruusalepp, 2008, Klopp, 2010). Forscher erhalten die Möglichkeit, ihre Daten bei PsychData langfristig zu archivieren und zur Nachnutzung bereitzustellen. Dies beinhaltet auch die Zitierfähigkeit der Daten, da diese über DOIs (Digital Object Identifiers) identifizierbar bleiben. Die bereitgestellten Daten können von Sekundärforschern für die wissenschaftliche Forschung und Lehre nachgenutzt werden. Sowohl die Datenbereitstellung durch den Produzenten als auch die Weitergabe an Datennutzer sind vertraglich geregelt. Abbildung 1 zeigt die Verortung der PsychData-Plattform im Forschungskontext. Neben der Datenarchivierung und -bereitstellung unterstützt PsychData Wissenschaftler bei der retrospektiven und forschungsbegleitenden Dokumentation ihrer Daten.



**Abbildung 1: PsychData im Forschungskontext**

## 2 Voraussetzung zum Data Sharing

Die Archivierung und Bereitstellung von Forschungsdaten erfordert eine entsprechende Aufbereitung der Daten. Neben den technischen Maßnahmen zur Datensicherung wie z.B. Speicherung in geräte- und systemunabhängigen Formaten, ist vor allem eine langfristige Interpretierbarkeit der Daten sicherzustellen. Ohne eine umfassende Dokumentation sind die Forschungsdaten inhaltsleer und somit für eine Nachnutzung unbrauchbar. Diese Dokumentation sollte optimalerweise anhand von internationalen Standards erfolgen, um Interoperabilität und Durchsuchbarkeit der Inhalte zu gewährleisten. Ebenso wichtig wie die

---

Dokumentation sind Kontrollen der Forschungsdaten auf Integrität, Vollständigkeit, Konsistenz und Anonymisierung.

Diese notwendige Dokumentation und auch Datenüberprüfung wird im Forschungsalltag häufig nachlässig durchgeführt. Doch gerade eine Dokumentation des Forschungsprozesses von Beginn an wird als „Best Practice“ im Datenmanagement angesehen (Van den Eynden et al., 2009). Eine retrospektive Aufbereitung der Daten ist meist zeit- und arbeitsaufwändiger, einzelne Fragen lassen sich im Nachhinein möglicherweise gar nicht mehr klären. Gründe für die Vernachlässigung einer gründlichen Dokumentation werden vor allem in der mangelnden zur Verfügung stehenden Zeit oder auch den fehlenden „incentives“ vermutet (Wagner, 2010). Eine weitere Ursache dafür, dass die Datendokumentation keine selbstverständliche Routine im Forschungsalltag darstellt, könnte in der fehlenden Verbreitung von Standards und Richtlinien zur Dokumentation liegen und auch im Fehlen geeigneter Werkzeuge und Plattformen, die eine möglichst effiziente Dokumentation der Forschungsdaten erlauben (vgl. Postle et al., 2002). Gängige Statistikpakete unterstützen den Forschenden hierbei nur sehr unzureichend.

### 3 Dokumentationsstandards im PsychData-Archiv

Das psychologische Archiv PsychData verwendet zur Beschreibung der zur Verfügung gestellten Forschungsdaten ein Set von Metadaten, das sowohl den Forschungskontext als auch die Variablen selbst umfassend dokumentiert. Diese Metadaten orientieren sich an den international anerkannten Dokumentationsstandards *Dublin Core*<sup>3</sup> mit seinen 15 Kernelementen und *DDI Version 2 (Data Documentation Initiative)*<sup>4</sup>. DDI ist ein XML-Format, das speziell für die Beschreibung von sozialwissenschaftlichen Daten entwickelt wurde und in vielen Projekten

---

<sup>3</sup> <http://dublincore.org/>, letzter Zugriff am 08.11.2010

<sup>4</sup> <http://www.ddialliance.org/>, letzter Zugriff am 08.11.2010

weltweit eingesetzt wird<sup>5</sup>. Im April 2008 ist Version 3 des DDI Standards erschienen, die im Gegensatz zu der vor allem auf die Variablenbeschreibung konzentrierte Version 2 versucht den gesamten *data life cycle* abzubilden. Mittelfristig ist geplant, die PsychData-Metadaten im DDI 3 Format zur Verfügung zu stellen. Das von PsychData verwendete Metadatenset ist mit beiden Dokumentationsstandards kompatibel, wurde aber entsprechend erweitert, um den Besonderheiten der psychologischen Forschung gerecht zu werden (vgl. Fahrenberg, 2009, Weichselgartner, 2008). Kennzeichnend für die Psychologie ist die Erforschung nicht direkt beobachtbarer innerpsychischer Prozesse. Für deren Messung sind Konstruktbildungen, Operationalisierungen und die Entwicklung von Erhebungsinstrumenten sowie deren Validierung notwendig und müssen dementsprechend auch ausreichend dokumentiert werden.

Mit den PsychData-Metadaten erfolgt daher zum einen eine genaue Studienbeschreibung anhand von bibliografischen Angaben (wie Autor, Titel etc.), einer inhaltlichen Studiendarstellung (z.B. zugrundeliegende Hypothesen) und insbesondere Angaben zum Datenerhebungsprozess (verwendete Testverfahren, Methodenbeschreibung, Stichprobenziehung etc.).

Zum anderen werden zum eindeutigen Verständnis der Forschungsdaten selbst die einzelnen Variablen anhand eines Kodebuchs beschrieben. Das von PsychData verwendete Kodebuch-Schema enthält die Elemente: Variablenname, inhaltliche Variablenbeschreibung, Itemtext (z.B. die gestellte Frage oder Instruktionsanweisung), Wertebereich der gültigen Werte, Wertebereich der fehlenden Werte, Bedeutung der gültigen Werte und Bedeutung der fehlenden Werte.

Die Erstellung einer kompletten Dokumentation anhand einer Studienbeschreibung und eines Kodebuchs bedeutet einen relativ zeitintensiven Arbeitsaufwand. Dieser kann durch entsprechende technische Unterstützung jedoch wesentlich verringert werden.

---

<sup>5</sup> Eine Liste einiger Projekte, die DDI zur Dokumentation ihrer Daten einsetzen, ist unter <http://www.ddialliance.org/ddi-at-work/projects> verfügbar (letzter Zugriff am 25.10.2010).

## 4 Ein Werkzeug zur forschungsbegleitenden Dokumentation

Im Rahmen des PsychData-Projekts wurde ein webbasiertes Dokumentationstool entwickelt, das es Forschern erleichtern soll, ihre Daten umfassend zu dokumentieren und auf Konsistenz zu überprüfen. Das Dokumentationstool ermöglicht die Datenbeschreibung anhand des PsychData-Metadatensets sowie das Hochladen und Archivieren der Forschungsdaten selbst. Die Nutzung des Tools kann sowohl retrospektiv als auch forschungsbegleitend erfolgen. Bei der Bearbeitung eines Kodebuchs und der zugehörigen Datenmatrix erfolgen automatische Fehlerkontrollen, durch die Mindeststandards der Datenaufbereitung gewährleistet werden.

Neben der Funktion eines reinen Dokumentationswerkzeugs für einzelne Forscher besteht auch die Möglichkeit, die Anwendung als eine Data Sharing Plattform zu verwenden. Einzelne Mitarbeiter können Mitgliedern aus ihrer Forschungsgruppe Zugriff auf ihre Datensätze gewähren. Ebenso kann die Anwendung bei der Betreuung von Diplomarbeiten oder Dissertationen eingesetzt werden, indem Studierende, die mit dem Tool arbeiten, ihren jeweiligen Betreuern den Zugriff auf ihre Daten ermöglichen. Hierfür steht eine Rechteverwaltung zur Verfügung, mit der die Zugriffsrechte für andere Personen auf die eigenen Daten festgelegt werden können.

Ein Anwendungsfeld dieses Werkzeugs sind Datensammlungen, wie sie im Rahmen von Diplom-/Masterarbeiten oder auch Dissertationen entstehen. Diese wurden bisher nicht langfristig archiviert und publiziert - Studierende sind zwar in der Regel von ihrem Lehrstuhl angehalten, ihre Daten auf einem Datenträger wie CD-ROM zusammen mit der Prüfungsarbeit abzugeben, im Gegensatz zu den eigentlichen Arbeiten, die häufig auf Dokumentenservern abgelegt werden, verschwinden die Daten dann aber meist in irgendwelchen Stahlschränken und sind für eine Nachnutzung nicht verfügbar. Abgesehen von den fehlenden Zugriffsmöglichkeiten sind die Beschreibungen der Daten nicht standardisiert, was ein Nachrechnen der Ergebnisse durch den/die Betreuer/in der Prüfungsarbeit bzw. eine Nutzung der Daten durch Dritte in einem anderen Auswertungskontext sehr schwierig macht. Hier kann das Dokumentationstool einen Beitrag leisten, um zum einen die Verfügbarkeit von Datensätzen zu gewährleisten, zum anderen aber auch die Lehrenden bei der Betreuung studentischer Arbeiten zu unterstützen. Darüber hinaus führt

der Einsatz eines solchen Tools in Empiriepraktika oder ähnlichen Veranstaltungen die Studierenden an die Grundsätze der guten wissenschaftlichen Praxis heran.

Das Dokumentationstool ist eine PHP/MySQL-Anwendung und stellt somit relativ geringe Anforderungen an die darunterliegende Infrastruktur. Aus Nutzersicht besteht die Software aus den Komponenten Metadaten-Formular, Kodebuch-Upload/-Eingabe, Datenmatrix-Upload/-Eingabe, einer Rechteverwaltung und Funktionen zum Datenexport.

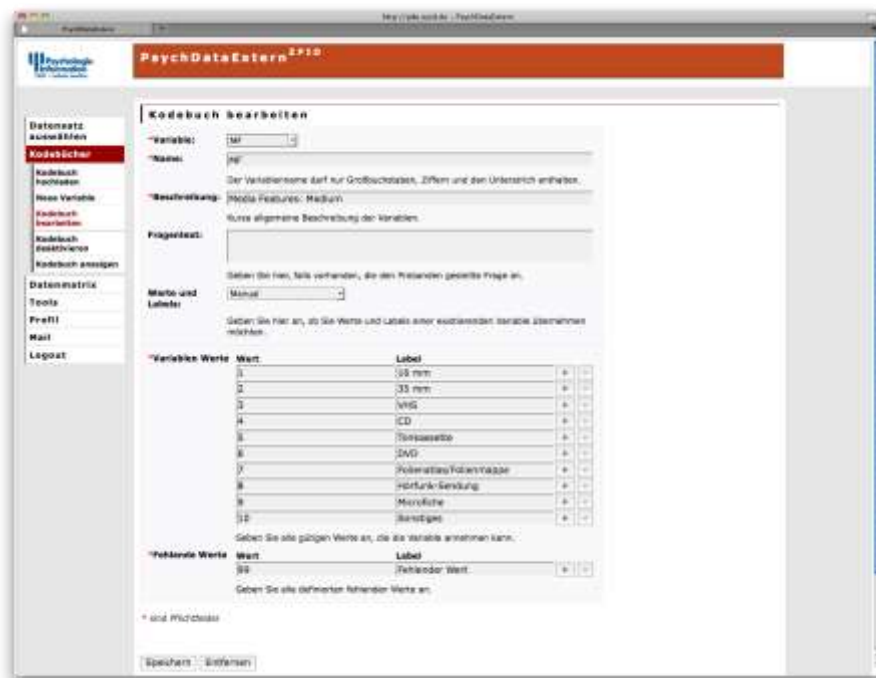
The screenshot shows a web browser window with the title 'PsychDataEtern 2.1.0'. The page has a red header bar with the application name. On the left, there is a navigation menu with items: 'Datenmatrix', 'Kodebücher', 'Datenmatrix', 'Tools', 'Profil', 'Mail', and 'Logout'. The main content area is titled 'Metadatenformular' and contains a form with several fields and text boxes. The fields include: 'Beteiligte Forscher' (with sub-fields for Name, Vorname, and Email), 'Arbeitstitel', 'Finanzielle Förderung', 'Verantwortlichkeit für die Datenerhebung', 'Copyright', 'Direkte Veröffentlichungen', 'Weiterführende Literatur', and 'Verwendete Quellen/Fürsorge'. To the right of these fields are explanatory text boxes. At the bottom of the form, there is a 'Speichern' button and a footer with version information: 'Datenmatrix (PsychDataEtern) Version: 2.1.0.0 (2012) Datenbank: 5.5.28a (32-bit)'.

Abbildung 2: Metadatenformular des Dokumentationstools

**Fehler! Verweisquelle konnte nicht gefunden werden.** zeigt das Metadatenformular der Anwendung, das in die Bereiche Basisdaten (z.B. beteiligte Forscher, Arbeitstitel, finanzielle Förderung), Informationen zur Studie (z.B. Forschungshypothesen, verwendete Erhebungsmethoden, Vorgehen, Zeiträume), Informationen zur Stichprobe und weitere Angaben zu den Daten (z.B. Anonymisierung und Informationen zur Kodierung der Daten) aufgeteilt ist. Diese Metadaten unterstützen später zum einen das

*Discovery* von Datensätzen, zum anderen machen sie die Entstehung der Daten nachvollziehbarer.

Ein wesentliches Element der Datendokumentation ist die Erstellung von Codebüchern. Die Anwendung bietet einen komfortablen Editor an, der die Nutzereingaben validiert und auf Fehler aufmerksam macht (siehe Abbildung 3). Alternativ zur direkten Eingabe in der Anwendung ist es auch möglich, vorbereitete Codebücher hochzuladen. Diese können beispielsweise in einem Texteditor erstellt werden. Geplant ist ein Werkzeug, das eine semiautomatische Erstellung der Codebücher aus SPSS heraus ermöglicht.



**Abbildung 3: Codebuch bearbeiten**

Ebenso wie die Codebücher werden auch die eigentlichen Daten direkt in der Datenbank verwaltet. Datenpunkte können entweder direkt eingegeben/bearbeitet werden oder in Form einer Textdatei mit Tab-separierten Werten hochgeladen werden. Die eingegebenen/hochgeladenen Werte werden gegen das Codebuch validiert, so dass Fehler direkt erkannt und korrigiert werden können.



Neben der Darstellung der Forschungsdaten in einer Datenmatrix steht auch eine Häufigkeitsverteilung über die einzelnen Variablen zur Verfügung (siehe Abbildung 4). Diese kann zur Kontrolle bereits exportierter Forschungsdaten dienen.

Als Exportfunktionen für Kodebuch und Datenmatrix existiert bisher die Ausgabe als Textdatei mit Tab-separierten Werten. Vorteil dieses Formats ist die Eignung als ein plattformunabhängiges Format. Geplant sind dennoch weitere Ausgabeformate wie Excel oder SPSS-Dateien.

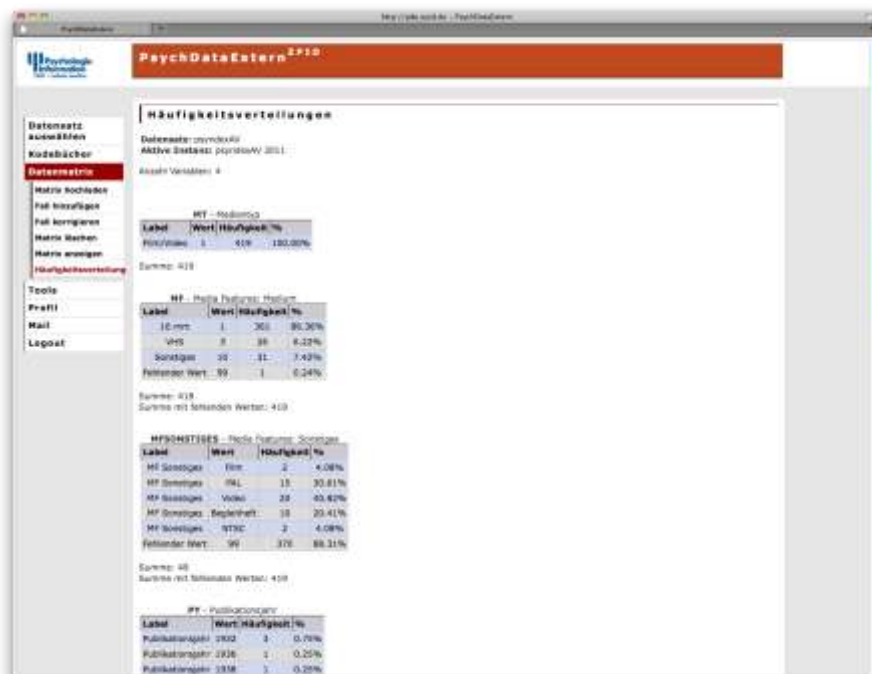


Abbildung 4: Häufigkeitsverteilung

Um die Anwendung als eine Data Sharing Plattform einsetzen zu können, beispielsweise im Kontext von Arbeitsgruppen oder studentischen Arbeiten, kann die Rechteverwaltung genutzt werden. Der Besitzer eines Datensatzes kann entscheiden, wem er welchen Zugriff auf einen Datensatz gewähren möchte. Zur Wahl stehen ein einfaches Leserecht, ein Lese- und Schreibrecht und darüber hinausgehend das zusätzliche *Grant*-Privileg, Zugriffsrechte an

weitere, im System registrierte Personen weiterzugeben. Abbildung 5 zeigt einen Screenshot der Rechteverwaltung.



Abbildung 5: Rechteverwaltung

## 5 Ausblick

Das psychologische Datenarchiv PsychData stellt ein Werkzeug bereit, das psychologische Forscher bei der forschungsbegleitenden Datendokumentation und -aufbereitung unterstützt. Außerdem Es wird eine sichere Archivierung der Forschungsdaten zusammen mit der zugehörigen Dokumentation ermöglicht. Dadurch können die Forschungsdaten sehr einfach für weitere Analysen bereitgestellt werden. Durch die Vorgabe von bewährten Dokumentationsstandards ergeben sich dabei mehrere Vorteile: Von Beginn der Forschungsarbeiten wird ein strukturiertes Vorgehen während des Forschungsprozess gefördert. Fehlerquellen wie ungültige Kodierungen werden durch die Validierungsroutinen sehr schnell erkannt und können direkt korrigiert werden.

Das Tool unterstützt noch nicht den gesamten Dokumentationsprozess - einige begleitende Arbeitsschritte können bisher nur außerhalb des Tools durchgeführt werden: Das Umwandeln von Statistik-Software-Formaten in tabseparierte Textdateien, das Erstellen einer Studiendokumentation und eines Kodebuchs, die Korrektur von Fehlern in Variablenbeschreibungen und Forschungsdaten setzen entsprechende Kenntnisse bzw. die Bereitschaft, sich diese anzueignen, voraus. Obwohl hier auch ein gewisser Vorteil gesehen werden kann, da sich der Nutzer aktiv mit Erfordernissen der Datendokumentation auseinandersetzen muss, kann es sich nur um einen zwischenzeitlichen Entwicklungsstand handeln. Die Erleichterung der Datendokumentation im Forschungsalltag ist eindeutiges Entwicklungsziel. Zur Erreichung dieses Ziels werden daher weitere Entwicklungen folgen wie z.B. die Erweiterung auf Ein- und Ausgabeformate gängiger Statistikprogramme oder das automatisierte Erstellen von Kodebüchern aus bereits vorhandenen (Teil-)Dokumentationen.

Der Mehrwert des Systems hat sich bereits in der Dokumentationsarbeit im psychologischen Datenarchiv PsychData gezeigt. Ein weiteres Anwendungsszenario ist die Nutzung des Dokumentationswerkzeugs als eine Art Data Sharing-Plattform für Forschungsprojekte.

## 6 Literaturverzeichnis

Bengel, J. & Wittmann, W.W. (1982). Bedeutung und Möglichkeiten von Sekundäranalysen in der psychologischen Forschung. *Psychologische Rundschau*, 33, 19-36.

Donders, F. C. (1868). Die Schnelligkeit psychischer Prozesse. *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin*, Berlin, 657-681.

Deutsche Forschungsgemeinschaft. (1998). *Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft — Weinheim: Wiley-VCH*

Erdfelder, E. (1994). Erzeugung und Verwendung empirischer Daten. In T. Herrmann & W. H. Tack (Hrsg.), *Enzyklopädie der Psychologie. Themenbereich B Methodologie und Methoden, Serie I Forschungsmethoden*

der Psychologie, Bd. 1 Methodologische Grundlagen der Psychologie (S. 47-97). Göttingen: Hogrefe.

Fahrenberg, J. (2009). Open Access – nur Texte oder auch Primärdaten? 05. Oktober 2009. Zugriff am 04.11.2010. Verfügbar unter [http://www.jochen-fahrenberg.de/fileadmin/openaces/Open\\_Access\\_Primaerdaten.pdf](http://www.jochen-fahrenberg.de/fileadmin/openaces/Open_Access_Primaerdaten.pdf)

Klopp, T. (2010). Open Data: Forscher sollen ihre Daten teilen. Zeit Online, 22. September 2010. Zugriff am 03.11.2010. Verfügbar unter <http://www.zeit.de/digital/internet/2010-09/data-sharing-open-access?page=1>

Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., Lautenschlager, M., Schindler, U., Sens, I. & Wächter, J. (2006). Data publication in the open access initiative. Data Science Journal, 5, 79-83. Zugriff am 05.11.2010. Verfügbar unter [http://www.jstage.jst.go.jp/article/dsj/5/0/5\\_79/article](http://www.jstage.jst.go.jp/article/dsj/5/0/5_79/article)

Postle, B. R., Shapiro, L. A. & Biesanz, J. C. (2002). On Having One's Data Shared. Journal of Cognitive Neuroscience, 14, 838-840.

Ruusalepp, R. (2008). Infrastructure, Planning and Data Curation. A comparative study of international approaches to enabling the sharing of research data. Version 1.6. 30. November 2008. Zugriff am 03.11.2010. Verfügbar unter [http://www.dcc.ac.uk/docs/publications/reports/Data\\_Sharing\\_Report.pdf](http://www.dcc.ac.uk/docs/publications/reports/Data_Sharing_Report.pdf)

Van den Eynden, V., Corti, L., Woollard, M. & Bishop, L. (2009). Managing and Sharing Data. A best practice guide for researchers. Essex: UK Data Archive. Zugriff am 04.11.2010. Verfügbar unter <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

Wagner, G.G. (2010). Forschungsdaten fallen nicht vom Himmel. Forschung und Lehre, 9, 650-651. Zugriff am 04.11.2010. Verfügbar unter <http://www.forschung-und-lehre.de/wordpress/?p=5562>

Weichselgartner, E. (2008). Fünf Jahre Primärdatenarchivierung in der Psychologie: Ein Erfahrungsbericht. In Ockenfeld, M. (Hrsg.), Verfügbarkeit von Information (S. 259-267). Frankfurt a. M.: DGI.

Wicherts, J.M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. American Psychologist, 61, 726-728