

Working Paper Series

www.ratswd.de

179

Disziplinspezifische Aspekte des Archivierens von Forschungsdaten am Beispiel der Psychologie

Erich Weichselgartner

Mai 2011

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienen empirischen Forschungsarbeiten, sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Disziplinspezifische Aspekte des Archivierens von Forschungsdaten am Beispiel der Psychologie

Erich Weichselgartner*

*Leibniz-Zentrum für Psychologische Information und
Dokumentation (ZPID) (wga@zpid.de)*

* Danksagung: Der Autor dankt Ina Dehnhard und Armin Günther vom PsychData-Team für die kritische Durchsicht des Manuskripts.

Abstract

Durch die immensen Fortschritte bei der Digitaltechnik sind nahezu beliebig hohe Speicherkapazitäten und weltumspannende Netzwerke selbstverständlich geworden. Der weitere Ausbau der Netzwerke zu Hochgeschwindigkeits-Grids ermöglicht neue Formen des wissenschaftlichen Arbeitens, die unter Begriffen wie e-Science oder virtuelle Forschungsumgebungen subsumiert werden. Digitale Objekte aller Art, seien es Texte, Bilder, Filme oder Tonaufzeichnungen, werden rund um den Globus archiviert und häufig kostenlos bereitgestellt. Fachdatenbanken wie PubMed oder Volltext-Repositorien wie arXiv sind Beispiele aus Medizin und Physik für den Technik-indizierten Wandel hin zu freier Verfügbarkeit wissenschaftlicher Information. Durch die Aufhebung der Speicherplatzbeschränkung geraten auch Bestandteile des wissenschaftlichen Forschungsprozesses in den Blickpunkt der Aufmerksamkeit, die man jahrzehntelang wenig beachtet hat: War es noch im 19. Jahrhundert in der Psychologie durchaus üblich, Forschungsdaten als Anhang von Publikationen abzudrucken, wurde dies im 20. Jahrhundert eingestellt. Erst im 21. Jahrhundert besinnt man sich wieder darauf, dass das Archivieren und Weitergeben von Forschungsdaten (*data sharing*) eine Vielzahl von Vorzügen hat. Die Möglichkeit zu Reanalysen und Metaanalysen einerseits, und die Vermeidung teurer Doppeluntersuchungen andererseits seien nur beispielhaft genannt.

Die erwähnten Vorzüge treten aber nur dann zutage, wenn die archivierten Forschungsdaten dauerhaft interpretierbar sind. Dazu müssen sie einerseits in einem Format gesichert werden, das robust gegenüber dem Technikwandel ist. Andererseits sind standardisierte Metadaten beizufügen, die sowohl die Variablen (Kodebuch) als auch die zugehörige Studie in toto (Kontext) umfassend beschreiben. Das psychologische Datenarchiv PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) setzt zur Festschreibung der Metadaten auf die Standards DC (Dublin Core metadata element set) und DDI (Data Documentation Initiative metadata specification for the social sciences). Zum Kontext gehören die Dokumentation des Erhebungsprozesses und der Erhebungsmethoden, sowie die Beschreibung der zugrunde liegenden Studie und ihres theoretischen Hintergrundes. Die Besonderheiten der psychologischen Forschung erfordern eine exakte und umfassende Dokumentation der Datenerhebung, weil es in der Psychologie kaum Normeinheiten wie den Archivmeter für die physikalische Längenmessung gibt. Deshalb ist es in der Psychologie auch so wichtig, die Forschungskultur im Fach dahingehend zu beeinflussen, dass eine forschungsbegleitende systematische Dokumentation von Datensätzen erfolgt. Spätere Rekonstruktionen binden ein Vielfaches der Ressourcen, als bei einer zeitnahen Dokumentation der Daten im Kontext der Erhebung benötigt worden wären. Da sich die psychologische Forschung mit dem menschlichen Erleben und Verhalten beschäftigt, ist der Anonymisierung der Daten und dem Datenschutz größte Aufmerksamkeit beizumessen.

Keywords: Psychologie, Forschungsdaten, Archivierung

Archivierung von Forschungsdaten

Grundsätzlich besteht über die Wissenschaftsdisziplinen hinweg weitgehend Einigkeit, dass Forschungsdaten erhalten und anderen zugänglich gemacht werden sollen. Eine bloße Archivierung ohne intendierte Weitergabe genügt zwar den Regeln zur Sicherung guter wissenschaftlicher Praxis von Forschungsförderern und Fachgesellschaften, doch erst die dauerhafte Archivierung in einer spezialisierten Institution und die bedingungslose Bereitstellung erzeugen einen Mehrwert für die Wissenschaftsgemeinschaft. Dies geht deutlich über den eingangs zitierten Minimalkonsens hinaus und bedarf in nahezu allen Disziplinen noch einer kopernikanischen Wende in der Forschungskultur, da das wissenschaftliche Belohnungssystem auf das Akkumulieren möglichst prestigeträchtiger Publikationen ausgelegt ist und die Ressourcen für die nachhaltige Bewahrung von Forschungsdaten schlicht fehlen. Es muss sich erst durchsetzen, dass bei der (Finanz-) Planung von Forschungsvorhaben auch an die nachhaltige Bewahrung der Forschungsdaten gedacht wird. Den Vorteil hätten alle, wie im Folgenden am Beispiel der Psychologie ausgeführt wird.

Vorzüge der Datenarchivierung am Beispiel der Psychologie

Ökonomie: Datenerhebungen sind mitunter sehr teuer und der geleistete Einsatz soll einen möglichst hohen Gewinn bringen, d.h. die Daten sollen möglichst umfassend ausgewertet werden. *Ethik:* Ein weiterer Grund, der gerade für Untersuchungen an Mensch und Tier gilt, ist die Vermeidung redundanter Datenerhebungen, da diese immer auch eine Belastung sein können. *Exklusivität:* Darüber hinaus sind manche Daten einmalig und können nicht repliziert werden (zum Beispiel solche, die unmittelbar nach Beendigung der deutschen Teilung erhoben wurden). *Exhaustivität:* Die Analyse der Datensätze unter neuen Fragestellungen und Perspektiven (neuen Forschungsansätzen), die Reanalyse der Datensätze mit anderen Analysemethoden, die vergleichende Analyse verschiedener Datensätze zur Ermittlung der Robustheit von Ergebnissen oder zur Ermittlung der Generalisierbarkeit empirischer Gesetze, die Aggregation von Testdaten zur Überprüfung und Optimierung von Testgütemerkmalen (auch zur Optimierung von Testnormen und zur Feststellung historischer Veränderungen

der Testnormen), die Ermittlung historischer Veränderungen und die präzisere Replikation von Untersuchungen, die auf der Basis der publizierten Forschungsberichte nur in grober Annäherung möglich wären. *Lehre, Kontrolle, Visibilität:* Daneben treten noch weitere Gründe wie die Ausbildung des wissenschaftlichen Nachwuchses mit realen Daten oder die Überprüfung publizierter Datenanalysen. Letzteres hat im Zusammenhang mit aufgedecktem wissenschaftlichem Fehlverhalten durchaus eine gewisse Brisanz. Nicht zuletzt soll die Bereitstellung von Primärdaten der Verstärkung des wissenschaftlichen Austauschs dienen und die Rezeption der Forschung durch die Fachöffentlichkeit unterstützen. Nach einer Untersuchung von Piwowar & Chapman (2008) haben Zeitschriften mit einer expliziten Richtlinie zur Datenbereitstellung einen höheren Impact Factor.

Disziplinäre Unterschiede des Forschungsdatenmanagements

Verschiedene Wissenschaftsorganisationen wie das International Council for Science (ICSU) und die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) machen sich für offenen Zugang zu Forschungsdaten stark (einen Überblick über entsprechende Open-Access-Initiativen und -Erklärungen bietet die Informationsplattform open-access.net). 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 (GenBank). Seit 1992 werden mit PANGAEA (Publishing Network for Geoscientific and Environmental Data) wissenschaftliche Daten aus allen Bereichen der Geowissenschaften und Biologie archiviert. 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-ökonomische Panel, SOEP). 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.

In der Psychologie gibt es von verschiedenen Seiten Empfehlungen zur Datenarchivierung, aber keinen Zwang. In Verlagsverträgen stimmen Autoren meistens zu, die einem Artikel zugrunde liegenden Primärdaten für eine bestimmte Frist (meist 5 – 10 Jahre) sicher aufzubewahren. Fachgesellschaften und Forschungsförderer äußern sich ähnlich. Die Deutsche Forschungsgemeinschaft (DFG) empfiehlt in ihrer Denkschrift *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*, 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). Seit 2003 ist in den ethischen Richtlinien des US-amerikanischen Psychologen-Verbandes American Psychological Association (APA) festgelegt, dass Forscher ihre Daten zu Veröffentlichungen anderen kompetenten Kollegen auf Wunsch zur Verfügung stellen sollen (APA, 2003). Die Deutsche Gesellschaft für Psychologie (DGPs) ist dieser Forderung in ihren auf die Forschung bezogenen ethischen Richtlinien gefolgt (DGPs, 2004).

In der Praxis scheinen die Selbstverpflichtungen aber nichts wert zu sein. Als Wicherts et al im Jahre 2005 versuchten, die Primärdaten von 249 Studien für eine Re-Analyse zur Verfügung gestellt zu bekommen, die ein Jahr vorher in APA-Zeitschriften veröffentlicht worden waren (für die also deren ethische Richtlinien gelten), erzielten sie nach intensiven Bemühungen einen enttäuschenden Rücklauf von 22%. Nach Meinung von Wicherts et al wäre es für Autoren zu aufwendig, von Dritten nachvollziehbare Metadaten und Codebücher bereitzustellen, noch dazu, ohne dafür materiell oder ideell entlohnt zu werden. Bereits 1999 stellte Azar fest: “There's little common ground on how, when and whether data should be shared [in Psychology]”. Der Entlohnungsaspekt wird neuerdings von Initiativen wie DataCite aufgegriffen, die Daten mittels persistenten Identifikatoren zitierfähig machen wollen, so dass sie analog zu Publikationen zur akademischen Reputation beitragen können (Lautenschlager & Sens, 2003). Mit dem Problem

der persönlichen Anerkennung für das Engagement im Forschungsdatenbereich setzen sich im Rahmen der RatSWD Working Paper Series Oellers, Siegers & Wagner (2011) auseinander.

Erkenntnisgewinnung in der Psychologie

Welche Forschungsdaten werden von der Psychologie generiert? Sind es qualitative oder quantitative Daten? Welche Datenmengen fallen an? Mit welchem Aufwand und nach welchen Standards sind die Daten zu dokumentieren, damit sie von anderen oder zukünftigen Forschern noch verstanden und genutzt werden können?

Die Psychologie im System der Wissenschaftsgliederung

Die moderne Psychologie wird typischerweise den Sozial-, Lebens- oder Naturwissenschaften zugeordnet. In vascoda, dem Portal für wissenschaftliche Information, ist die Psychologie in den Clustern „Medizin und Biowissenschaften“ sowie „Rechts-, Wirtschafts- und Sozialwissenschaften“ vertreten. Die internationale Suchmaschine Microsoft Academic Search ordnet die Psychologie den Lebenswissenschaften zu (life science). Im sog. Frascati-Handbuch (2006) der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) findet man die Psychologie bei den Sozialwissenschaften. In den USA, dem sowohl in wissenschaftlicher als auch anwendungspraktischer Hinsicht wichtigsten Land für die Psychologie, wird die Psychologie häufig den Kognitionswissenschaften (cognitive science), Neurowissenschaften (neural science) oder Verhaltenswissenschaften (behavioral science) zugeordnet. An diesen Zuordnungen erkennt man, dass sich die moderne Psychologie eher an dem gesellschafts- und naturwissenschaftlichen Forschungsansatz orientiert als an dem geisteswissenschaftlichen.

Datenerhebung und Messung in der Psychologie

Die Psychologie ist die Wissenschaft vom menschlichen Erleben und Verhalten. Sie ist eine empirische Wissenschaft, d.h. sie bedient sich bei der Erkenntnisgewinnung der Erfahrung. Ihre Forschungsthemen, wie zum Beispiel das Denken, das Fühlen oder das Wollen, sind nicht direkt beobachtbar, sondern müssen über

psychologiespezifische Forschungsmethoden erschlossen werden. Die Datenerhebung in der Psychologie basiert hauptsächlich auf systematischen und kontrollierten Beobachtungen im Feld oder im Labor. Als Königsweg für die Erkenntnisgewinnung gilt für viele das Experiment, weil es die Vorzüge Inter-subjektivität, Präzision und Wiederholbarkeit aufweist. Problemstellung, Forschungsweg und Ergebnis sind beim Experiment mehr als bei anderen Methoden *öffentlich* (Janke, 1969) und damit auch den Regeln guter wissenschaftlicher Praxis unterwerfbar (DFG, 1998). Das Experiment, die Verhaltensstichprobe oder auch der Test finden unter standardisierten und kontrollierten Bedingungen statt, d.h. sie werden bei allen Personen zu allen Zeitpunkten in der gleichen Weise durchgeführt. Das Experiment ist objektiv, wenn verschiedene Auswerter aufgrund derselben Daten zu übereinstimmenden Ergebnissen kommen. Die Regeln zur Administration, Auswertung und Interpretation werden eindeutig festgelegt. Tack (1994) fordert von den empirischen Methoden, dass sie Aussagen liefern, die auf *nachvollziehbaren* Beobachtungsergebnissen beruhen.

Wie heute in den empirischen Wissenschaften üblich, werden in der Psychologie die beobachteten Phänomene (Beispiel: Anzahl pro Zeiteinheit gelöster Aufgaben) in Zahlen überführt (*Scoring*). Die registrierten Phänomene werden so quantifiziert und können mit Digitalrechnern leichter weiterverarbeitet und – zu meist statistisch – analysiert werden. Bei den so kodierten Zahlenmatrizen als Kombination aus kontrollierten (unabhängigen) und unkontrollierten (abhängigen) Variablen handelt es sich um die Forschungsdaten im engeren Sinn. Da für die Datenanalyse in der Psychologie häufig das Statistik-Programm SPSS¹ herangezogen wird, besteht ein Großteil des Datenmaterials, die an das psychologische Datenarchiv PsychData übergeben wird, aus SPSS-Dateien (Archivierungsformat von PsychData ist aber ASCII, siehe unten).

Ein großes Problem der Psychologie besteht darin, dass das Vorhandensein von Zahlen zu Rechenoperationen verleitet (z.B. Subtraktion oder Division), obwohl die Zahlen eigentlich nur als Benennungen dienen (wie die Trikotnummern bei Fußballern); Mausfeld (1994) spricht in diesem Zusammenhang von „Zahlzeichen“. Man betrachte als Beispiel drei Personen, deren Körpergröße und Intelligenz man misst. Person A sei 160 cm groß, Person B 170 cm und Person C 180

1 Statistical Package for the Social Sciences

cm. Dann lässt sich eine Rangordnung aufstellen, die man für die empirische Welt ausdrücken kann als $A \prec B \prec C$ (\prec bedeutet „sichtbar kleiner“); führt man das empirische Relativ in ein numerisches Relativ über, kann man die Beziehung ausdrücken als: $A < B < C$ ($<$ bedeutet „mathematisch kleiner“). Im Falle von Körpergröße liegt mehr als eine Rangskala vor, nämlich mindestens eine Intervallskala, wie man per einfacher Beobachtung mit Hilfe eines Referenzmaßes feststellen kann: Person B ist 10 cm kleiner als Person C, und Person A ist 10 cm kleiner als Person B. Die numerischen Beziehung ($180 \text{ cm} - 170 \text{ cm} = 170 \text{ cm} - 160 \text{ cm}$) hat eine empirische Entsprechung dergestalt, dass Person B um genau das gleiche Stück Maßband kleiner ist als Person C, wie Person A kleiner ist als Person B.

Die Frage ist nun, ob für Intelligenz das gleiche gilt. Person A habe einen IQ von 90, Person B von 100 und Person C von 110. Dann ist Person B um 10 Punkte weniger intelligent als Person C und Person A ist um 10 Punkte weniger intelligent als Person B. Hat dann die numerischen Beziehung ($110 \text{ IQ-Punkte} - 100 \text{ IQ-Punkte} = 100 \text{ IQ-Punkte} - 90 \text{ IQ-Punkte}$) eine empirische Entsprechung dergestalt, dass Person B um genau den gleichen Wert weniger intelligent ist als Person C, wie Person A weniger intelligent ist als Person B? Liegt also auch für Intelligenz eine Intervallskala vor (wo Differenzen bedeutsam sind), oder vielleicht nur eine Rangreihe (welche die empirische Aussage rechtfertigt: „C ist intelligenter als B, und B ist intelligenter als A“, ohne dass die Abstände auf der Dimension *Intelligenz* zwischen den drei getesteten Personen gleich sind)?

Im Falle der Körpergröße ist der Längenunterschied unmittelbar anhand eines geeichten Stück Maßbands *augenfällig*, während man im Falle der Intelligenz mehr tun muss als nur hinschauen. Die Schwierigkeit, nicht direkt-beobachtbare Merkmale (z.B. Intelligenz, Motivation, Zufriedenheit) zu messen und das entsprechende Skalenniveau zu bestimmen, ist eine der wesentlichen Herausforderungen der Psychologie. Stärker als andere empirische Disziplinen, bei denen man bei der Messung auf Standards wie beispielsweise das Urmeter in Paris zurückgreifen kann, muss sich die Psychologie mit der Theorie des Messens auseinandersetzen (vgl. Abschnitt Dokumentation unten). Zur Einführung seien als grundlegende Werke zur Skalierung und zur Messtheorie empfohlen: Luce et al. (1971), Pfanzagl (1968), Roberts (1979) und Stevens (1946).

Geisteswissenschaftliche Psychologie

Von dem eben geschilderten objektiv-quantitativen Weg zur Erkenntnis unterscheidet sich der qualitativ-subjektive Forschungsansatz, wie er in der geisteswissenschaftlichen Psychologie verwendet wird: Zum Einsatz kommen hier deutende Methoden wie Biographik, Exploration und Inhaltsanalyse. Das vorliegende Datenmaterial, zum Beispiel Audiomitschnitte von Interviews oder Videomitschnitte von Gruppendiskussionen, muss aufwendig transkribiert werden (bei Videoaufzeichnungen zum Beispiel verbale und nonverbale Äußerungen), damit es anschließend interpretiert werden kann (Hermeneutik).

Forschungsdaten in der Psychologie

In der Psychologie fallen also zwei Arten von Forschungsdaten an: qualitative Daten und quantitative Daten. Zu ersteren gehören nicht-standardisierte Beobachtungen, Interviews und selbstproduzierte Texte wie z.B. Tagebücher. Zu letzteren gehören standardisierte Beobachtungen, Befragungen, Ergebnisse von Experimenten und psychophysiologische (EEG, EKG, EDA, fMRT) oder psychoneuroendokrinologische Messwerte (Hormondatenmessung mittels Blutabnahme oder Speichelprobe). Kritiker monieren an ersteren Unexaktheit, Subjektivität und Verzerrungen, an letzteren die mangelnde inhaltliche Substanz der Fragestellungen.

Für die Archivierung von psychologischen Forschungsdaten ergibt sich aus den dargelegten Besonderheiten der psychologischen Forschung die Notwendigkeit einer sehr exakten und umfassenden – und damit sehr aufwendigen – Dokumentation der Daten und des Datenerhebungskontexts. Wenn sich ein Physiker mit einer Messreihe von Temperaturdaten befasst, dann ist das Instrument, mit dem die Messungen vorgenommen wurden, in der Regel irrelevant, weil verschiedene (geeichte) Messinstrumente das gleiche Messen. Selbst wenn die Werte in unterschiedlichen Einheiten vorliegen (Celsius, Fahrenheit, Kelvin, Réaumur), können sie mit simplen Formeln umgerechnet werden. Wenn sich ein Psychologe der Intelligenz von Probanden befasst, so ist die Angabe des Messinstruments (des speziellen Testverfahrens) und der zugehörigen Eichstichprobe unabdingbar. In der Psychologie muss also sehr aufwendig dokumentiert werden, und zwar sowohl die Variablen (Kodebücher) als auch der Erhebungskontext (Metadaten)².

2 Auch die Kodebücher sind Metadaten, die Unterscheidung wird mitunter zur engeren Spezifizierung gebraucht.

Datenarchivierung in der Psychologie

Die Archivierung von Forschungsdaten hat keine Tradition in der Psychologie. Zwar war es im 19. Jahrhundert durchaus üblich, bei Publikationen im Anhang auch die Forschungsdaten (Primärdaten, Rohdaten) abzudrucken (z.B. Donders, 1868), aber diese Praxis kam im 20. Jahrhundert wieder aus der Mode und wäre angesichts der schieren Datenmenge im vorelektronischen Zeitalter auch nicht praktikabel gewesen. Heute findet man psychologische Forschungsdaten, oder Daten, die auch für die Psychologie interessant sind, in Deutschland häufig im sozial-, wirtschafts- und erziehungswissenschaftlichen Kontext. Beispiele für entsprechende Archive sind – in dieser Reihenfolge – das GESIS Datenarchiv für Sozialwissenschaften (DAS), das Sozio-oekonomische Panel (SOEP) und das Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen (IQB). Die Datenservicezentren (DSZ) und Forschungsdatenzentren (FDZ) der empirischen Sozial-, Wirtschafts- und Verhaltenswissenschaften haben sich im Ständigen Ausschusses Forschungsdaten–Infrastruktur (FDI) des Rates für Sozial- und Wirtschaftsdaten (RatSWD) zusammengeschlossen. Die Psychologie-Suchmaschine PsychSpider des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) erfasst derzeit neun Deutsche und ein Schweizer Archiv³.

Auch im sprachwissenschaftlichen und medizinischen Kontext der Psychologie findet man im In- und Ausland Archive. In Nijmegen (Niederlande) unterhält das Max-Planck-Institut für *Psycholinguistik* das weltweit größte *Archiv bedrohter Sprachen*; in Pittsburgh (USA) wird das *Child Language Data Exchange System* betrieben; die University of Western Australia pflegt die *Medical Research Council Psycholinguistic Database* und in Ann Arbor (USA) gibt es das *Substance Abuse and Mental Health Data Archive*. Einen USA-lastigen Überblick über Datensätze und Datenrepositorien in der Psychologie gibt die American Psychological Association auf ihren Webseiten (siehe <http://www.apa.org/research/responsible/data-links.aspx>).

Auch die Rückkehr zur eingangs erwähnten Praxis, die Forschungsdaten zusammen mit dem Forschungsaufsatz zu publizieren, ist neuerdings zu beobachten. So wurde das *Archive of Norms, Stimuli, and Data* der Psychonomic Society (USA) aufgelöst und von Springer in das elektronische Angebot der Zeitschrift

³ http://www.zpid.de/PsychSpider_Hilfe.php?tip=25

Behavior Research Methods integriert. Beispiel: Ruft man mit Hilfe des Digital Object Identifiers (DOI) den Aufsatz von Clark & Paivio (2004) auf (DOI: 10.3758/BF03195584), erhält man einen Link auf den zugehörigen Datensatz, den man als ZIP-Datei ohne weitere Restriktionen, aber mit der Auflage, die Quelle zu zitieren, herunterladen kann.

Psychologisches Datenarchiv PsychData

Das bisher einzige dezidiert für die Psychologie eingerichtete Datenarchiv für quantitative Daten mit breitem Sammlungsfokus ist PsychData, das vom Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) in Trier betrieben wird. In PsychData sollen bedeutsame Primärdatenbestände aus der psychologischen Forschung vorgehalten werden. Fachgebiete sind die gesamte Psychologie mit Schwerpunkt auf Datensätzen aus den Bereichen der Klinischen, der Entwicklungs-, der Pädagogischen, der Geronto- sowie der Arbeits- und Organisationspsychologie, die im Rahmen von Längsschnittstudien, groß angelegten Umfragestudien und Testentwicklungen erhoben wurden. Die Arbeiten an PsychData wurden im Jahre 2002 mit DFG-Unterstützung aufgenommen (Montada & Weichselgartner, 2002; Krampen & Weichselgartner, 2005). Im Gegensatz zu manchen Forderungen einer verpflichtenden Datenweitergabe⁴ setzt PsychData auf Freiwilligkeit, qualitative Selektion und vertraglich fixierte Nutzung. Die Datengeber können für die Bereitstellung mit oder ohne Auflage einer Karenzzeit optieren. In der Praxis haben sich bisher rund 10% der Datengeber für eine Karenzzeit von 5 Jahren entschieden. Wegen der Besonderheiten der personenbezogenen Daten in der Psychologie ist vor Weitergabe der Daten durch das ZPID immer ein schriftlicher Vertragsschluss notwendig (siehe <http://www.psychdata.de/>). Die Daten werden nicht über das Netz, sondern nur auf revisionssicheren CD-ROMs geliefert. Lediglich die Metadaten (Beschreibung der Datensätze) sind uneingeschränkt frei verfügbar.

Archivierungsobjekte von PsychData sind Metadaten, Codebücher und Forschungsdaten in Form von Zahlenmatrizen, wie sie in der Psychologie zur weiteren Verwendung durchgängig erstellt werden (siehe Abschnitt Daten-

4 Das US-amerikanische Bundesgesundheitsamt (National Institutes of Health) verlangt seit 2003, dass die Primärdaten geförderter Projekte im Sinne des Open Access spätestens nach einer Frist von sechs Monaten öffentlich bereitgestellt werden müssen (NIH, 2003).

erhebung und Messung in der Psychologie). Forschungsdaten lassen sich in nicht-proprietären Formaten speichern und bieten zudem den Vorteil, einer Sekundärnutzung unmittelbar zugeführt werden zu können. Das Archivierungsformat entspricht dabei nur in wenigen Fällen dem Format der überlassenen Datensätze, die in der Regel im Speicherformat der verwendeten Auswertungs-Software (in der Psychologie zumeist SPSS) geliefert werden. Diese Datensätze sind daher von PsychData-Mitarbeitern entsprechend der Archivierungsstandards zu prüfen, zu bearbeiten und zu wandeln.

In der Forschungspraxis erfolgt die Verwahrung der Daten nicht immer in optimaler Form. Die verwendeten Formate und Speichermedien genügen oftmals nicht dem Ziel einer langfristigen Archivierung. Bei unzureichender Archivierung können die Daten – sei es durch System- oder Formatwechsel, technische Probleme oder menschliches Versagen – unwiederbringlich verloren gehen. Beispiele von innerhalb und außerhalb der Psychologie gibt es genug. Sogar die NASA musste den Verlust der Daten von 1,2 Mio. Magnetbändern bekannt geben (Hoegner, 1999). Die Archivierung dient dem Schutz der Datensätze vor Datenverlust, -alterung und -manipulation. Bei PsychData wird dies in der technischen Umsetzung durch quelloffene Software (u.a. Unix, MySQL, Apache, PHP), das Zusammenspiel von mehreren räumlich getrennten Servern, definierte Backups auf magnetischen und optischen Medien, Prüfsummen und einer abgestuften Zugriffskontrolle erreicht. Als Datenformat wird grundsätzlich das plattform- und herstellerunabhängige Format ASCII eingesetzt. Daten, die in proprietären Formaten angeliefert werden, werden konvertiert. Die Migration auf jeweils aktuelle Systeme ist vorgesehen.

Dokumentation

Zur langfristigen Interpretierbarkeit von Datenbeständen ist die Dokumentation der archivierten Inhalte unabdingbar. Diese umschließt die Beschreibung der Zahlwerte selbst (Kodebuch) sowie die Dokumentation des Kontextes der Datenerhebung, also etwa des Messinstrumentes, der Experimentalsituation oder auch der Untersuchungsstichprobe (Metadaten). So fordert beispielsweise Knobloch (2004) für die Teilchenphysik explizit die umfassende Beschreibung des Kontextes der Datengenerierung: „data are useless without additional documentation, metadata, and software“. In den Lebenswissenschaften und insbesondere in der

Psychologie kommt hinzu, dass kaum Messinstrumente vorliegen, die an einem objektiven Standard geeicht werden können (vgl. Abschnitt Datenerhebung und Messung in der Psychologie). Zugleich sind die Untersuchungsobjekte im Regelfall von erheblich höherer Komplexität als die meisten physikalischen Systeme. Obschon der mit der Datendokumentation verbundene Aufwand damit in der Psychologie gegenüber den physikalischen Wissenschaften deutlich ansteigt, werden gerade aufgrund dieser Merkmale datenbeschreibende Arbeitsschritte besonders dringlich. Die fehlende eindeutige Verbindung von Konstrukt und Operationalisierung macht es im Gegenzug möglich, dass die Daten auch unter anderen als den vom Primärforscher bearbeiteten Fragestellungen genutzt werden können (Bien, 2002). Durch die Bereitstellung einheitlicher Standards zur Daten- und Untersuchungsdokumentation (Kodebuch, Metadaten) kann die Interpretierbarkeit der Datenbestände auch bei längerfristigen Nutzungszeiträumen, bei wechselnder personaler Ausstattung des Forscherteams oder bei Sekundärnutzung garantiert werden. Leider kommt es durch Personalfluktuationen in wissenschaftlichen Forschungsprojekten häufig dazu, dass vorhandene Datensätze nicht mehr interpretierbar sind, da nur unmittelbar mit dem Datensatz beschäftigte Personen – und diese oft auch nur für einen begrenzten Zeitraum – alle mit den Daten verknüpften Aspekte nachvollziehen können.

Metadaten und Kodebücher

Das PsychData-Metadatenschema lehnt sich an die Dublin Core Metadata Initiative (DCMI, <http://dublincore.org/>) und an die internationale Data Documentation Initiative (DDI, <http://www.ddialliance.org/>) an. DCMI stammt aus dem bibliothekarischen Kontext, während DDI für die Dokumentation sozialwissenschaftlicher Studien entwickelt wurde.⁵ Die Anlehnung an DDI gewährleistet die Interoperabilität zu anderen Archiven und bewerkstelligt durch eine entsprechende Adaptation die Beschreibung psychologischer Daten bei gleichzeitig möglichst geringer Belastung der Datengeber. Aufgabe der Metadaten ist es nicht nur, den Kontext der Datenerhebung festzuhalten, sondern die Primärdaten für potentielle Nutzer (im Internet) besser auffindbar zu machen. Sie liefern Grundinformationen über einen Datensatz, wie zum Beispiel Angaben über Autor, Titel

5 DDI umfasst auch Metadaten im Sinne der Studiendokumentation und des Kontextes der Datenerhebung. Explizit sind auch die 15 Kernelemente von Dublin Core enthalten.

oder Zeitpunkt der Datenerhebung, Die Metadaten der in PsychData archivierten Studien und entsprechende Suchmöglichkeiten stehen der Fachöffentlichkeit seit 2004 zur Verfügung.

Zur Datendokumentation dient in der Nomenklatur der Sozialwissenschaften das Kodebuch (Codebuch). Es ermöglicht die effiziente, erfolgreiche und fehlerfreie Nutzung der Primärdaten. Im Kodebuch findet man die Beschreibung der Variablen und der dazugehörigen Werte (Namen und Ausprägungen). Zur Verdeutlichung folgendes Beispiel:

In einer Untersuchung wurde das Geschlecht der Versuchspersonen erfragt. Wenn Sie nun die Daten dieser Untersuchung einsehen oder weiterverwenden wollen, so finden Sie im Kodebuch umfassende, gleichzeitig aber so knapp wie möglich gehaltene Informationen zu der Variable „Geschlecht“. Neben dem Variablennamen und der Variablenbeschreibung findet sich z.B. das Fragebogenitem, wie es im Original vorlag: „Bitte nennen sie Ihr Geschlecht“. Außerdem werden die Bedeutungen der einzelnen Werte erklärt, nämlich dass der Forscher das Geschlecht *männlich* mit einer 1, das Geschlecht *weiblich* mit einer 2 und fehlende Werte mit einer 9 kodierte. Auch Zusatzinformationen, wie dass es sich um eine Filtervariable handelt, sind im Kodebuch festgehalten.

Retrospektive vs studienbegleitende Datendokumentation

Bei der retrospektiven Datendokumentation handelt es sich um die ausführliche Beschreibung von Primärdaten bereits abgeschlossener Studien. Sie ist ein wesentliches und zeitintensives Arbeitsfeld für das PsychData-Projekt-Team. Einen Überblick über die Arbeitsschritte vom Eingang eines Datensatzes bis zu dessen Präsentation findet sich in Abbildung 1 (vgl. Weichselgartner, 2008).

Der idealtypische Workflow zeigt, dass die Bearbeitung der Datensätze Prüfungsschritte, datensatzmodifizierende Schritte und verschiedene Erstellungsschritte beinhaltet. Der Aufwand der direkt datensatzmodifizierenden Schritte und der verschiedenen Erstellungsschritte hängt dabei unmittelbar vom Umfang sowie der Qualität des eingereichten Datenmaterials ab. Insgesamt zielen die in Abbildung 1 spezifizierten Schritte darauf ab, bestehende Inkonsistenzen zwischen Daten, Variablenbeschreibung, Erhebungsinstrument und den Angaben des Primärforschers zur Studie zu beseitigen und somit die Präsentation unmittelbar verwertbarer, umfassend dokumentierter Datenbestände zu ermöglichen. Zwar tragen

informationstechnologische Entwicklungen zu einer steten Optimierung der Arbeitsabläufe bei, so dass die retrospektive Primärdatendokumentation zunehmend effizienter gestaltet werden kann, trotzdem muss ein vergleichsweise großer Aufwand von mehreren Arbeitstagen bis -wochen je Studie einkalkuliert werden.

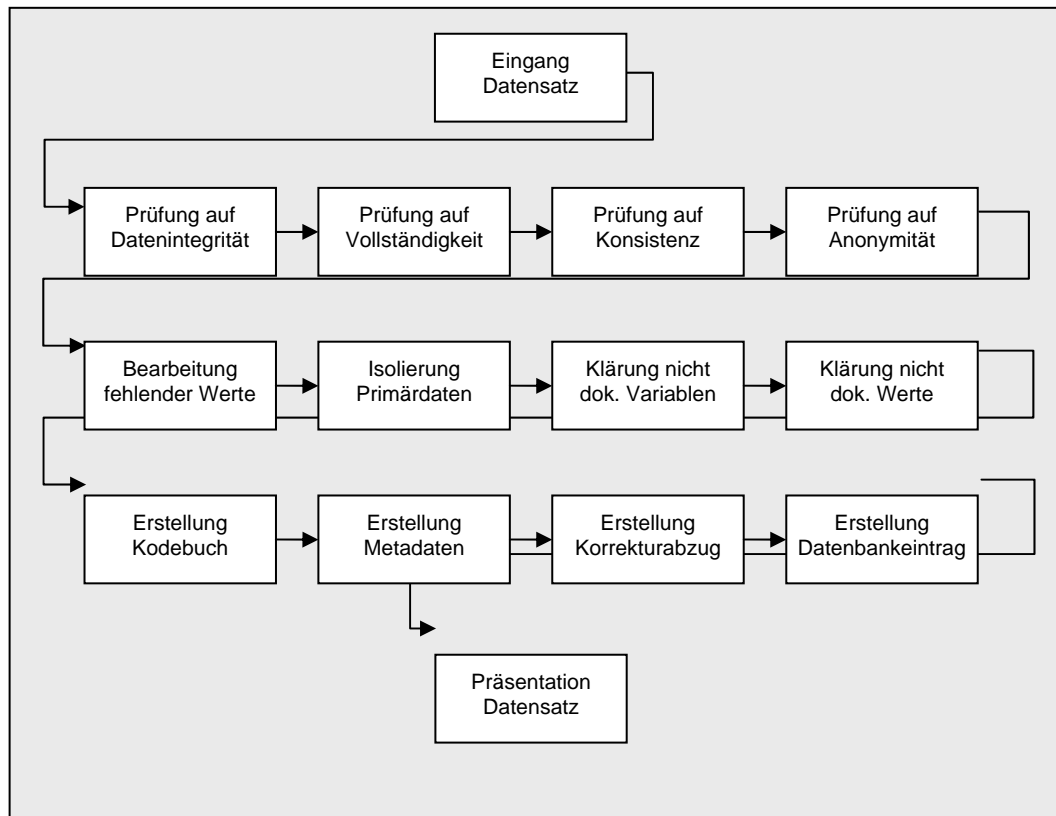


Abbildung 1. Beispielhafter Workflow der retrospektiven Forschungsdatendokumentation und -archivierung in PsychData

Die *studienbegleitende Datendokumentation* sieht die technische und instruktio- nale Unterstützung von Primärforschenden bezüglich der Dokumentation und Archivierung bereits in der Planungs- oder Erhebungsphase vor. Mit dem for- schungsbegleitenden Ansatz wird den Problemen der retrospektiven Aufbereitung – Verlust oder Leseprobleme von Daten, mangelnde Rekonstruierbarkeit von Variablen- und Wertelabeln etc. – wirkungsvoll begegnet. Darüber hinaus trägt dieser Ansatz dazu bei, den Primärforschungsprozess effizient zu strukturieren und zu optimieren. Durch dieses Angebot soll eine Standardisierung des for- schungsbegleitenden Dokumentationsprozesses und damit eine Gewöhnung der Forschenden an stringente Dokumentationsverfahren erreicht werden. Durch die Bereitstellung entsprechender Werkzeuge und Manuale soll der Aufwand einer

forschungsbegleitenden Dokumentation soweit reduziert werden, dass er den Aufwand einer retrospektiv notwendig werdenden Dokumentation – etwa bei der Abfassung von Forschungsberichten oder der Übergabe an Sekundärforscher – deutlich unterschreitet.

Für die instruktionale Unterstützung steht das PsychData-Handbuch zum Download bereit⁶. Ziel dieses Handbuchs ist es, Datengeber bei Ihrem Dokumentationsprozess zu unterstützen und so die Eingabe in das Datenarchiv vorzubereiten.

Die technische Unterstützung geschieht mit Hilfe eines web-basierten Tools zur forschungsbegleitenden Dokumentation. Dieses soll empirisch tätigen Psychologen und Psychologinnen ebenfalls bei der Dokumentation ihrer Forschungsarbeiten als Unterstützung dienen, und zwar unabhängig davon, ob sie eine Datenarchivierung in PsychData in Betracht ziehen oder nicht. Bei einer gewünschten Datenübergabe an das Archiv lässt sich dann die gesamte Kommunikation mit den Forschenden über eine einheitliche Oberfläche abwickeln, was zu einer erheblichen Optimierung des Workflows beiträgt. Eine Kontrolle bestimmter Prüfschritte ist dann bereits bei der Bearbeitung durch den Primärforscher möglich. So können z.B. bei der Studienbeschreibung Hilfestellung und Rückmeldung zu obligatorischen Dokumentationsfeldern gegeben werden. Bei der Erstellung der Variablenbeschreibung durch ein Kodebuch können sowohl das Kodebuch selbst auf Konsistenz geprüft werden als auch anschließend die Übereinstimmung mit den Datenmatrizen. Auf diese Weise lassen sich an früher Stelle im Forschungsprozess Eingabefehler, nicht dokumentierte Variablen, ungeklärte fehlende Werte, etc. verhindern. Idealziel ist ein in den Forschungsprozess integriertes *data-lifecycle* Management. Die erste Version eines Dokumentationstools für externe Anwender wurde im Mai 2010 online bereitgestellt. Für eine Nutzung ist eine einfache Registrierung notwendig. Bisher haben sich 10 externe Nutzer/innen im Dokumentationstool registriert.

6 http://psychdata.zpid.de/downloads/PsychData-Handbuch_mrz2010.pdf

Integration in wissenschaftliche Nachweissysteme

Wie oben ausgeführt, dient die Datendokumentation auch dazu, die Daten für potentiell interessierte Forscher mühelos auffindbar zu machen. Um dies auf möglichst breiter Front zu unterstützen, werden mehrere Wege beschritten.

Suchmaschinenoptimierung. Selbst im wissenschaftlichen Kontext werden heute häufig verbreitete horizontale Suchmaschinen wie Bing oder Google eingesetzt. Die Datendokumentation ist deshalb so für das Web aufzubereiten, dass die Suchmaschinen-Crawler barrierefreien Zugang haben. Darüber hinaus betreibt das ZPID mit dem PsychSpider eine vertikale Suchmaschine für die Psychologie, die eine spezielle Kollektion (Cluster) für Forschungsdaten enthält. Dadurch werden bei der Recherche irrelevante Treffer von vorneherein ausgeschlossen.

Fachdatenbanken. Um potentiell interessierte Forscher auf das Vorhandensein von Datensätzen aufmerksam zu machen, sollen bei der Literatur- oder Personenrecherche entsprechende Hinweise angezeigt werden. Das ZPID hat deshalb die Literaturdatenbank PSYINDEX und das Datenarchiv PsychData miteinander verknüpft: Bei einer Suche in PSYINDEX erhält der Nutzer/in eine Zusatzinformation, wenn Forschungsdaten zur gesuchten Literatur in PsychData vorhanden sind. Durch die Verlinkung kann dann direkt die Erhebungsbeschreibung zu den Forschungsdaten aufgerufen werden. Umgekehrt wird auch von den zu einem Forschungsdatensatz gehörigen Literaturangaben direkt zu dem in PSYINDEX erfassten Dokument verlinkt. Eine anloge Verlinkung ist für die Personendatenbank PsychAuthors in Vorbereitung.

Digital Object Identifier. Persistente Identifikatoren dienen der dauerhaften Kennzeichnung von Datensätzen und anderen digitalen Objekten und garantieren deren Auffindbarkeit im Internet. Mit der Einführung von persistenten Identifikatoren für die bereitgestellten PsychData-Datensätze wird zudem die Zitierbarkeit dieser Primärdaten ermöglicht. Als persistente Identifikatoren bieten DOIs (Digital Object Identifier) den Vorteil, das mit ihrer Registrierung auch die Angabe von Metadaten verbunden ist. Diese Metadaten enthalten Informationen, die auf den vorgesehenen Anwendungsbereich abgestimmt sind. Aus diesem Grund wird für die archivierten PsychData-Datensätze eine Registrierung von DOIs bei der Registrierungsagentur da|ra für sozialwissenschaftliche Datensätze vorgenommen. Betreiber von da|ra ist das Leibniz-Institut für Sozialwissenschaft-

ten (GESIS), welches Mitglied im DataCite-Verbund ist. Mit da|ra werden die Voraussetzungen für eine dauerhafte Identifizierung, Sicherung, Lokalisierung und schließlich eine verlässliche Zitierbarkeit von Forschungsdaten geschaffen.

Praxis

Es wurde bereits angemerkt, dass die Archivierung von Forschungsdaten keine Tradition in der Psychologie hat. Bedingt durch das mit PsychData vorhandene Serviceangebot und den allgemeinen Trend zur freien Verfügbarkeit wissenschaftlicher Information wandelt sich die Forschungskultur im Fach allmählich. Die Nutzung in beide Richtungen, d.h. die Übergabe von Daten an das Archiv und der Bezug von Daten aus dem Archiv, steigt kontinuierlich an. Für die Dokumentation der eingereichten Datensätze sind die Ressourcen derzeit ausgeschöpft. Das liegt sowohl an dem hohen Aufwand für die Dokumentation psychologischer Forschungsdaten per se, als auch an dem hohen Aufwand für die retrospektive Dokumentation von Längsschnittstudien, die vor vielen Jahrzehnten begonnen wurden. Hier müssen ständig Unklarheiten und Inkonsistenzen recherchiert werden, wodurch es immer wieder zu Unterbrechungen und zeitaufwendigen Rückkoppelungsschleifen zwischen Datengebern, assoziierten Primärforschern und den Dokumentaren kommt. Beispiele hierfür sind die *Longitudinalstudie zur Genese individueller Kompetenzen* (Weinert, 1999), von der bislang ein Viertel der insgesamt 323 Datensätze dokumentiert und kontrolliert werden konnte, und die *Genetisch orientierte Lebensspannenstudie zur differentiellen Entwicklung* (Weinert & Geppert, 1996), die bereits im Jahre 1937 begonnen wurde. Weitere große Längsschnittstudien, die zur Bearbeitung an das Archiv übergeben wurden, sind die *Rostocker Längsschnittstudie* (Meyer-Probst & Reis, 1999), die die Entwicklung von 294 Rostocker Bürgern seit ihrer Geburt 1970/1971 verfolgt, und die Längsschnittstudie zu *Jugendentwicklung und Drogen* mit 7 Datensätzen und insgesamt 6.222.206 Datenpunkten (Kastner & Silbereisen, 1987). Die bisher größte Probandenzahl (21.189) hat die Studie *Fahren unter Alkohol in Deutschland* (Krüger & Vollrath, 2004), die bisher größte Variablenzahl (2.982) die Kölner Scheidungsstudie 1990-1996 (Schmidt-Denter & Beelmann, 2004).

Die in der psychologischen Diagnostik eingesetzten Testverfahren (kurz Tests) müssen aufwendig konstruiert und an geeigneten (repräsentativen) Stichproben normiert werden. Viele werden von Verlagen kommerziell vertrieben. Umso

erfreulicher ist es, dass die Inhaber der Publikationsrechte der Veröffentlichung einer Reihe von Testnormierungsdatensätzen in PsychData zugestimmt haben.

Die Nachnutzung der Forschungsdaten findet in unterschiedlichen disziplinären Kontexten statt (nicht nur in der Psychologie, sondern beispielsweise auch in den Ingenieurwissenschaften, Sportwissenschaften und Sozial- und Wirtschaftswissenschaften), und zwar nicht nur für die Forschung, sondern auch für die Lehre und für akademische Qualifikationsarbeiten (Diplomarbeiten, Dissertationen).

Infrastruktur für die Fachgemeinschaft

Mit einem zentralen Angebot zur Unterstützung von Primärforschenden bei der Datendokumentation und Datenarchivierung können kostspielige Mehrfachentwicklungen vermieden und der Fachöffentlichkeit unter Einsparung von Ressourcen eine zukunftsfähige Infrastruktur überlassen werden. Durch die Übernahme konzeptueller und informationstechnologischer Grundlagenarbeit durch das ZPID sowie die Gewährung praktischer Hilfestellung bei der Datenaufbereitung und -dokumentation ist für die Forschenden die Konzentration auf den inhaltlichen Beitrag möglich.

Das ZPID gibt die entwickelten Systeme auf Wunsch weiter und unterstützt damit Forschungsinstitute bei der institutsinternen Datendokumentation und Datenarchivierung. Entsprechende Kooperationen mit einer norddeutschen und einer österreichischen Universität zeigen, wie die PsychData-Administrationsumgebung außerhalb des ZPID für Primärforscher eingesetzt werden kann.

Literatur

- American Psychological Association. (2003). Ethical Principles of Psychologists and Code of Conduct. — URL: <http://www.apa.org/ethics/code2002.html> [08.08.2008]
- Bien, W. (2002). Primärerhebung oder Reanalyse? Überlegungen und Beispiele für die Optimierung des Ressourceneinsatzes durch Sekundäranalysen. In H. Merckens & J. Zinnecker (Hrsg.), *Jahrbuch Jugendforschung 2002* (S. 173-189). Opladen: Leske und Budrich.
- Clark, J. M. & Paivio, A. (2004). Extensions of the Paivio, Yuille and Madigan (1968) norms. *Behavior Research Methods*, 36, 371-383. DOI: 10.3758/BF03195584
<<http://www.springerlink.com/content/t867n76u00184933/>>
- Deutsche Forschungsgemeinschaft. (1998). Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Weinheim: Wiley-VCH.
- Deutsche Gesellschaft für Psychologie. (2004). Revision der auf die Forschung bezogenen ethischen Richtlinien. — URL: <http://www.dgps.de/dgps/aufgaben/ethikr12004.pdf> [08.08.2008]
- Donders, F. C. (1868). Die Schnelligkeit psychischer Prozesse. *Archiv für Anatomie, Physiologie und wissenschaftliche Medizin*, Berlin, 657-681.
- Kastner, P. & Silbereisen, R. K. (1987). Jugendentwicklung und Drogen: Eine prospektive Längsschnittstudie als wissenschaftliche Begründung präventiver Intervention. In W. Specht (Hrsg.), *Die gefährliche Straße: Jugendkonflikte und Stadtteilarbeit* (S. 227-253). Bielefeld: Boellert.
- Key Perspectives (2010). Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study. Digital Curation Centre. [online] Abrufbar unter <http://www.dcc.ac.uk/scarp> (Zugriff am 17.02.2011).
- Knobloch, J. (2004). Disciplinary Case Study 1. Physical Sciences – Particle Physics at CERN and elsewhere. — URL: www.erpanet.org/events/2003/lisbon/presentations/CernCodataLisbon.pdf [27.05.2011]
- Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971), *Foundations of measurement*, Vol. I: Additive and polynomial representations. New York: Academic Press.
- Janke, W. (1969). Das Experiment in der Psychologie. In Salber, W. u. a.: *Methoden der Psychologie und Pädagogik. (Enzyklopädie der geisteswissenschaftlichen Arbeitsmethoden. 7. Lieferung)*. München: R. Oldenbourg Verlag.
- Krampe, G. & Weichselgartner, E. (2005). Dokumentation und Archivierung von Rohdatensätzen aus der psychologischen Forschung. DFG Sachbeihilfe GZ: 554 922 (1) Uni Trier BIB44 TRuv 01-02.
- Lautenschlager, M. & Sens, I. (2003). Konzept zur Zitierfähigkeit wissenschaftlicher Primärdaten. *Information Wissenschaft & Praxis*, 54, 463-466.
- Mausfeld, R. (1994). Von Zahlzeichen zu Skalen. In T. Herrmann & W.H. Tack (Hrsg.), *Enzyklopädie der Psychologie, Themengebiet B, Serie I, Band 1*. Göttingen: Hogrefe.
- Meyer-Probst, B. & Reis, O. (1999). Von der Geburt bis 25 Jahre – Rostocker Längsschnittstudie. *Z. Kindheit und Entwicklung* 8, 1, 59-68.
- Montada, L. & Weichselgartner, E. (2002). Dokumentation und Archivierung von Rohdatensätzen aus der psychologischen Forschung. DFG Sachbeihilfe GZ: 554 922 (1) Uni Trier BIB44 TRuv 01-01.
- National Institutes of Health. (2003). Final NIH Statement on Sharing Research Data. — URL: <http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html> [08.08.2008]
- OECD (2006). *Frascati Manual – Proposed Standard Practice for Surveys on Research and Experimental Development*. Paris: OECD. Annex 1 „Fields of science and technology classifications“ abgerufen am 9.5.2011 von: <http://www.oecd.org/dataoecd/36/44/38235147.pdf>
- Oellers, C., Siegers, R. & Wagner, G.. (2011). Zum Problem der Anerkennung der Arbeit an Forschungsdaten: Welchen Beitrag können Persistente Identifikatoren für Forschungsdaten und Autoren leisten? Working Paper. RatSWD Working Paper Series.
- Pfanzagl, J. (1968). *Theory of Measurement*. Würzburg/Wien: Physica-Verlag.
- Piwowar H.A. & Chapman, W.W. (2008). A review of journal policies for sharing research data. *Nature Precedings*: hdl:10101/npre.2008.1700.1.
- RIN (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] Abrufbar unter <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-annex.pdf> (Zugriff am 17.02.2011).
- Roberts, F. S. (1979). *Measurement Theory*. Reading, MA: Addison Wesley.
- Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103 (2684), 677–680.
- Tack, W. H. (1994). Ziele und Aufgaben einer allgemeinen Methodenlehre der Psychologie. In Th. Herrmann & W. H. Tack (Hrsg.), *Enzyklopädie der Psychologie (Themenbereich B, Serie I, Band 1)* (S. 2-44). Göttingen: Hogrefe.

- 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.
- Weinert, F. E. & Geppert, U. (Hrsg.). (1996). *Genetisch orientierte Lebensspannenstudie zur differentiellen Entwicklung (GOLD)*. Report Nr. 1: Planung der Studie. München: Max-Planck-Institut für psychologische Forschung.
- Weinert, F. E. (Hrsg.). (1999). *Individual Development from 3 to 12. Findings from the Munich Longitudinal Study*. Cambridge: Cambridge University Press.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The Poor Availability of Psychological Research Data for Reanalysis. *American Psychologist*, 61, 726-28.