

Datenmanagement und Datenbereitstellung

Wo kann ich meine Daten ablegen und wie muss ich sie für eine potenzielle Nachnutzung aufbereiten?

Martin Kerwer, Ina Dehnhard, Armin Günther, Ronny Bölter, Erich Weichselgartner

Leibniz-Zentrum für Psychologische Information und Dokumentation

Projekt DataWiz

Projektleitung: Erich Weichselgartner

Projektmitarbeiter: Ronny Bölter, Ina Dehnhard, Armin Günther, Martin Kerwer

14:00 – 14:45	Forschungsdatenmanagement und Bereitstellung von Daten
14:45 – 15:00	<i>Pause</i>
15:00 – 15:45	Datenaufbereitung
15:45 – 16:00	<i>Pause</i>
16:00 – 16:30	Einführung DataWiz
16:30 – 17:30	Testung DataWiz
17:30 – 18:00	Rückmeldung, Diskussion & Abschluss

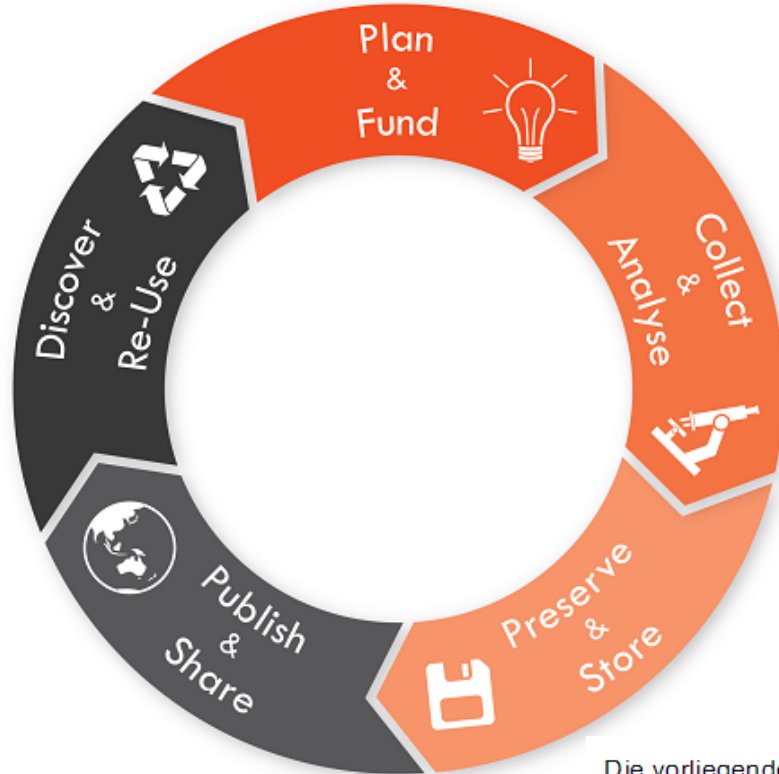
Forschungsdatenmanagement bezeichnet den Prozess, der alle Methoden und Verfahren umfasst, die zur Sicherung der langfristigen Nutzbarkeit von Forschungsdaten angewendet werden: die Generierung, die Bearbeitung, die Anreicherung, die Archivierung und die Veröffentlichung.

<http://www.forschungsdaten.org/index.php/Forschungsdatenmanagement>

Dabei sind die verschiedenen Dimensionen der Arbeit mit Daten jeweils und in ihrem Zusammenspiel zu berücksichtigen. Diese sind insbesondere

- *rechtliche und ethische Dimensionen (Urheberrecht, Datenschutz)*
- *technische Dimensionen (Soft- und Hardware)*
- *dokumentarische Dimensionen (Metadaten).*

<http://www.forschungsdaten-bildung.de/datenmanagement>



URL: <https://library.sydney.edu.au/research/data-management/research-data-management.html#>

DGPs | Deutsche Gesellschaft
für Psychologie

DOI: <http://dx.doi.org/10.1026/0033-3042/a000341>

Die vorliegende Konkretisierung der DFG-Leitlinien verfolgt das Ziel, eine Balance zwischen diesen unterschiedlichen Interessen zu realisieren. Sie

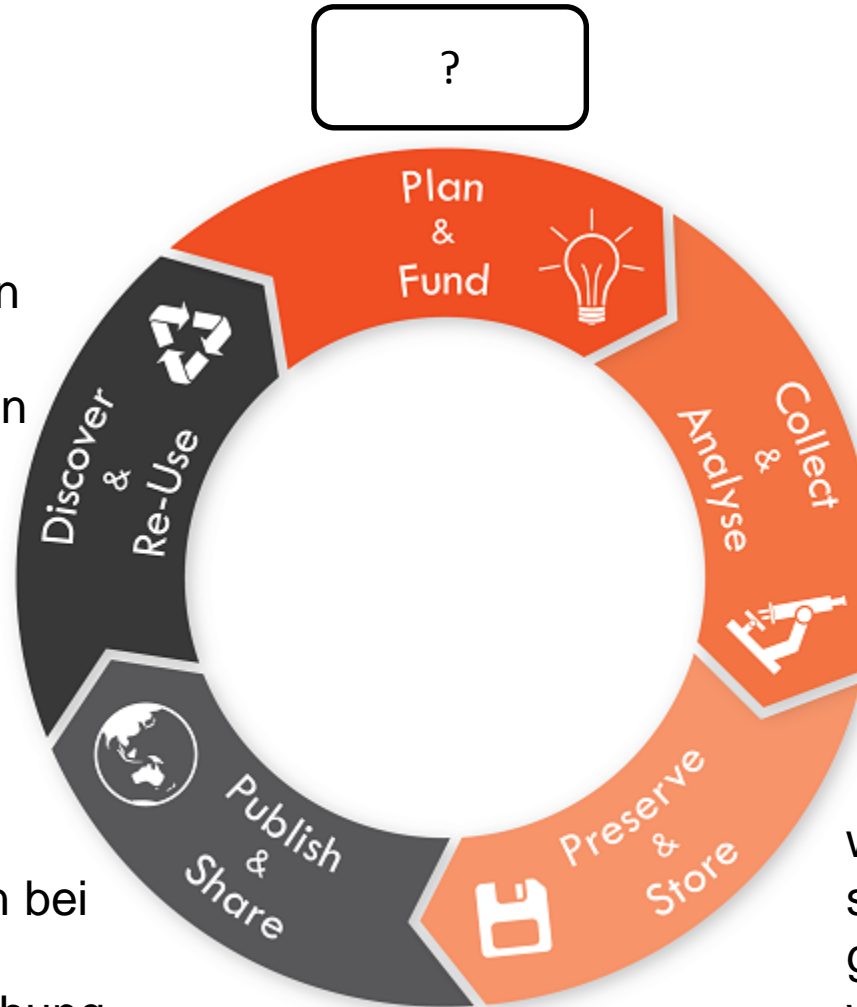
- thematisiert die Bedeutung eines nachhaltigen Managements von Forschungsdaten,
- definiert, was „Primärdaten“ sind und wie sie gespeichert werden sollen,
- definiert Standards und eventuelle Einschränkungen bei der Veröffentlichung von Forschungsdaten und
- definiert die Rechte und Pflichten von „Datenbereitstellenden“ und „Datennachnutzenden“.

Schönbrodt, Gollwitzer, & Abele-Brehm, 2016

DGPs Empfehlungen zum FDM

Rechte und Pflichten von
Datenbereitstellenden
und Datennachnutzenden

Standards und
Einschränkungen bei
der
Datenveröffentlichung



was „Primärdaten“
sind und wie sie
gespeichert
werden sollen

Und warum sollte ich mich darum schon frühzeitig kümmern?

Aus den DGPs-Empfehlungen (S. 5):

Einwilligungserklärungen und Ethikanträge sollten daher so angepasst werden, dass sie kompatibel mit den vorliegenden Empfehlungen sind.

Auch wenn in den DGPs Empfehlungen nur ein Endzustand spezifiziert wird, muss man frühzeitig im Projektverlauf die Weichen dafür stellen, dass er verwirklicht werden kann.

- Forschungsfrage(n)/Ziel(e)
 - Förderung beantragen
 - Ethikantrag
 - Einwilligungserklärungen
 - Ein- und Ausschlusskriterien definieren
- Materialien & Methoden
 - Bereits existierende Datenbestände (z.B. NEPS, Collaborative Psychiatric Epidemiology Surveys)
 - Dokumentations- und Metadatenstandards?
 - Nutzung von etablierten Messinstrumenten oder Eigenentwicklung?
 - Software
- Workflowentwicklung
 - Präregistrierung von Studien
 - Manuale oder Standard-Operating-Procedures
 - Datenanalyse (Ein- und Ausschlusskriterien)



- **BDSG**
 - Besondere Arten personenbezogener Daten beachten
 - Anonymisierung ist auch eine Datenverarbeitung
 - Wenn Einwilligung zur Weitergabe vorliegt, sollten auch personenbezogene Daten teilbar sein → Nachnutzung jenseits von Nachfolgeprojekten.

Use of the information I provide for this project only

I understand my personal details such as phone number and address will not be revealed to people outside the project.

☐ ☐

Use of the information I provide beyond this project

I agree for the data I provide to be archived at the UK Data Archive.²

☐ ☐

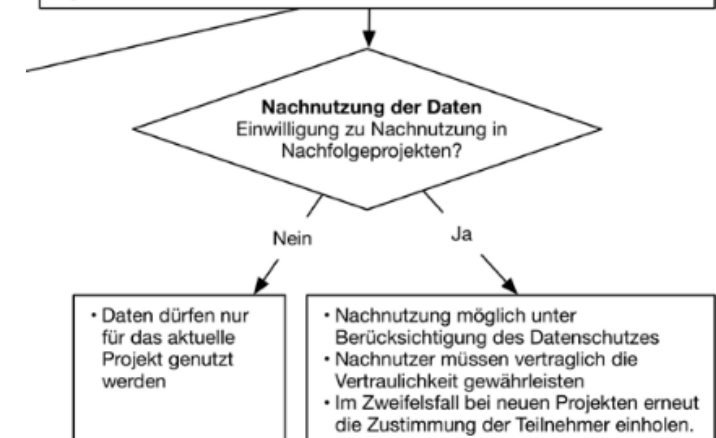
URL: <https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms>

Folgende Elemente in Teilnehmerinformation:

- Einwilligung zur Datennutzung im aktuellen Projekt einholen
 - fakultativ (s.u.): Einwilligung zur Nachnutzung in weiteren Projekten einholen, mit möglichst spezifischer Beschreibung potentieller Nachnutzungen (falls Beschreibung zu unspezifisch kann die Einwilligung ungültig sein)
- Information, dass zur wissenschaftlichen Qualitätssicherung die Daten 10 Jahre gespeichert bleiben sollen
 - fakultativ (s.u.): Einwilligung zur Speicherung der Daten auch für länger als 10 Jahre einholen
- Information über Recht, die Einwilligung in die weitere Nutzung der eigenen Daten jederzeit zu widerrufen und die Löschung zu verlangen
- Datenbesitzende Stelle und Verantwortlichkeit für personenbezogene Daten dauerhaft und klar regeln (An wen wende ich mich im Fall eines Widerrufs?)

Qualitätssicherung bei personenbezogenen Daten:

- Nachprüfung der Ergebnisse im aktuellen Projekt möglich, indem externen Forschern vertraulicher Zugang zu den personenbezogenen Daten gegeben wird; Nachprüfer müssen vertraglich die Vertraulichkeit gewährleisten



Schönbrodt, Gollwitzer, & Abele-Brehm, 2016



The OKCupid dataset: A very large public dataset of dating site users



Content removed

This content has been removed

OSF

[Explore](#)

[Contact](#)

[Help/Support](#)

[Guides](#)

[FAQ](#)

Center for Open Science

[Home](#)

[Reproducibility Project: Psychology](#)

[Reproducibility Project: Cancer Biology](#)

[TOP Guidelines](#)

[Donate](#)

URL: <https://osf.io/p9ixw/>

Copyright © 2011-2016 [Center for Open Science](#) | [Terms of Use](#) | [Privacy Policy](#)

q30;Would you like to receive pain during sex?;Yes

q32;Group sex (with 3 or more people);I have tried it.;It seriously interests me.;I have little or no interest;;32816;0;;sex/intimacy

q33;Do you like kissing in public?;Yes;No;;;23004;0;;sex/intimacy

q34;Are you more horny or more lonely?;Horny;Lonely;;;23592;N;;sex/intimacy

q35;Regardless of future plans

q36;Would you ever consider an open marriage? That means you can sleep with other people.;Yes;No;;;16649;0;;sex/intimacy

Beispiel Nutzung bestehender Fragebögen



- Copyright der Verfahren liegt bei Verlagen (Psychologiespezifisch)
- Copyright auch auf die Manuale (Skalenberechnungen, Normdaten) und Formulierungen
- Transparente Dokumentation ohne Urheberrechtsverletzung

Einsatzbereich

Der FAKT-II dient der adaptiven Ermittlung der individuellen Konzentrationsfähigkeit. Die Anwendungsbereiche des FAKT-II liegen in der Eignungsdiagnostik, der Arbeits- und Betriebspsychologie, Klinischen Psychologie, Schulpsychologie, Pädagogischen Psychologie, Verkehrspsychologie, Sportpsychologie, im neuro- und pharmakopsychologischen Bereich, in der Psychiatrie sowie in der Grundlagenforschung.

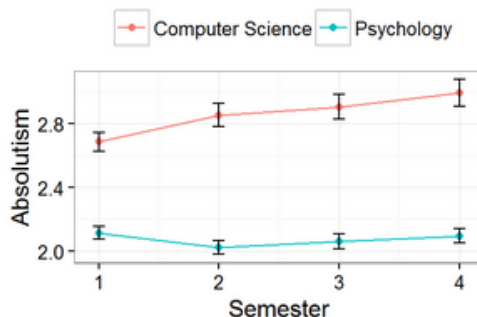
URL: <https://www.testzentrale.de/shop/frankfurter-adaptiver-konzentrationsleistungs-test-ii-69356.html>

- Forschungsdaten erheben
 - Forschungsdaten in digitale Formate übertragen und dokumentieren
 - Forschungsdaten analysieren
 - Entblindung vornehmen
-
- Ggf. Anpassung geplanter Workflows:
 - Analysen
 - Auswahlkriterien
 - Vorgehen bei der Datenerhebung
 - Dokumentation der Änderungen!



- Qualitätssicherung und Data Cleaning
 - Doppelte Dateneingabe (in Planung festlegen)
 - Prüfung der Wertebereiche (über Syntax)
 - Kodierung: Fehlende Werte nicht in den Bereich gültiger Werte legen
- Dokumentation besonderer Ereignisse während des Experiments (Störungen, Entblindung, Protokollabweichungen, Abweichungen vom Manual)

- Berechnung von Skalen oder Erstellung von Grafiken nachvollziehbar dokumentieren (z.B. rmarkdown)



```
#Funktionen und Packages laden
source("Functions_20150707.R")
library(xlsx)
library(plyr)
library(multilevel)
library(psychometric)
```

```
df5 <- rbind(t(describe(df1[df1$group == "Psychologie",], na.rm = T)$mean),
t(describe(df1[df1$group == "Informatik",], na.rm = T)$mean),
t(describe(df2[df2$group == "Psychologie",], na.rm = T)$mean),
t(describe(df2[df2$group == "Informatik",], na.rm = T)$mean),
t(describe(df3[df3$group == "Psychologie",], na.rm = T)$mean),
t(describe(df3[df3$group == "Informatik",], na.rm = T)$mean),
t(describe(df4[df4$group == "Psychologie",], na.rm = T)$mean),
t(describe(df4[df4$group == "Informatik",], na.rm = T)$mean))
```

```
df6 <- rbind(t(describe(df1[df1$group == "Psychologie",], na.rm = T)$se),
t(describe(df1[df1$group == "Informatik",], na.rm = T)$se),
t(describe(df2[df2$group == "Psychologie",], na.rm = T)$se),
t(describe(df2[df2$group == "Informatik",], na.rm = T)$se),
t(describe(df3[df3$group == "Psychologie",], na.rm = T)$se),
t(describe(df3[df3$group == "Informatik",], na.rm = T)$se),
t(describe(df4[df4$group == "Psychologie",], na.rm = T)$se),
t(describe(df4[df4$group == "Informatik",], na.rm = T)$se))
```

Plots for Article

```
levels(df$group)[1] <- "Computer Science"
levels(df$group)[2] <- "Psychology"

p1 <- ggplot(data=df, aes(x=time, y=EPI_A, group=group)) + theme_bw() +
  geom_line(aes(color=group)) +
  geom_point(aes(color=group)) +
  geom_errorbar(aes(ymin=EPI_A - EPI_A_se, ymax=EPI_A + EPI_A_se, width=.1) +
  theme(legend.position="top") + scale_x_continuous(labels=c("1", "2", "3", "4"), name = "Semester") + scale_y_co
ntinuous(name = "Absolutism") + theme(legend.title=element_blank())
##theme(axis.line = element_line(color = 'black'))
```

- Forschungsdaten zur Nachprüfbarkeit sichern
 - Zusammenstellung des Archivierungspakets (Forschungsdaten, Dokumentation etc.)
 - Überführung von Daten in Archivierungsformate
 - (Finale) Aufbereitung der Daten
 - Nachnutzbarkeit der bereitgestellten Daten sicherstellen
 - Kriterienkatalog Anhang A: DGPs Empfehlungen
- Forschungsdaten für die Nachnutzung in einem Repository bereitstellen
 - Auswahl der Daten
 - DGPs Empfehlungen: Publikationstypen
 - Wahl eines geeigneten Repository
 - Datengebervertrag/Lizenzen festlegen
 - Übergabe der Daten



1 Grundlegende Kriterien:

- Wirtschaftliche & ideologische Unabhängigkeit sowie wissenschaftliche Professionalität*
- Ausrichtung: Fächerübergreifend vs. disziplinspezifisch
- Rechtliche Aspekte (z.B. länderspezifische Jurisdiktion) *
- Aufnahmekriterien für Daten
- Datennutzung einschränkbar (z.B. Embargo) *
- Persistente Identifikatoren (z.B. DOI) *
- Kosten *

*DGPs Empfehlungen

2 Technische Aspekte:

- Umgang mit verschiedenen Datenformaten *
- Langfristige Speicherung der Daten*
- Ablage zusätzlicher Daten wie Instrumente, Software, Analyseskripte, etc. *
- Versionierung von Datensätzen

3 Inhaltliche Aspekte:

- Dokumentation auf Studienebene *
- Dokumentation auf Variablenebene *
- Zitationspflicht * und /oder Ko-Autorenschaft *
- Qualitätssicherung: Beratung, Unterstützung & Anleitung*

***DGPs Empfehlungen**

Suchmaschinen, Linklisten, u.a. :

- Re3data.org
- DataCite.org
- PsychLinker.de
- RatSWD.de
- ...

PsychLinker
Forschungsdaten



RatSWD.
Rat für Sozial- und
WirtschaftsDaten
Akkreditierte Datenzentren

- **Nimmt das Repository Daten „von Externen“ an?**
 - Institutionelle Repositorien (z.B. heiDATA)
 - Projektspezifische Repositorien (z.B. SOEP, NEPS, PISA)
- **Gibt es spezielle Aufnahmekriterien für Daten:**
 - Disziplin- oder themenspezifisch
 - Quantitativ oder qualitativ
 - spezifische Datentypen (z.B. fMRI, EEG,...)
- **Sind spezielle Aufnahmekriterien für meine Daten relevant?**
 - Auffindbarkeit
 - datenschutzrechtliche Aspekte
 - besondere Erfordernisse an Dokumentation

Spezialisierte Repositorien

Qualitative Daten

Qualiservice qualiservice.org

FDZ Bildung fdz-bildung.de



Spezifische Datentypen

OpenfMRI openfmri.org

(NITRC -IR nitrc.org)



Spezielles Fachgebiet:

Dryad datadryad.org



Auswahl eines Repositoriums

PsychData psychdata.de

Forschungsdatenzentrum für die Psychologie

GESIS gesis.org

DAS – Datenarchiv für Sozialwissenschaften

datorium – Self-Deposit Daten-Repository

Zenodo zenodo.org

Repository (EU)

Open Science Framework osf.io

Open Source Software Project

Auswahl eines Repositoriums

DANS EASY easy.dans.knaw.nl

Data Archiving and Networked Service

Harvard Dataverse <https://dataverse.harvard.edu/>

Open Source Software Project (z.B. heiDATA)

figshare <https://figshare.com/>

ICPSR

„Klassisch“ (für Mitglieder) <http://www.icpsr.umich.edu/icpsrweb/>

openICPSR Self-Deposit <https://www.openicpsr.org/openicpsr/>

Allgemeine Kriterien - Vergleich

	PsychData	GESIS datorium	Zenodo	OSF
Wiss. Unabhängigkeit Professionalität	ZPID Leibniz-Institut	GESIS Leibniz-Institut	CERN	COS
Ausrichtung	Psychologie	Social Science	Alle	Alle
Jurisdiktion	Deutschland	Deutschland	Schweiz	USA
Aufnahmekriterien	-psychologische Daten -peer-reviewed pub.	Keine	Keine	Keine
Datennutzung einschränkbar	✓	✓	✓	✗
Persistente Identifikatoren	✓	✓	✓	✓
Kostenfreie Datenübergabe	✓	✓	✓	✓
Kostenfreie Datennutzung	✓	✓	✓	✓

Allgemeine Kriterien - Vergleich

	DANS	Harvard Dataverse	figshare	openICPSR
Wiss. Unabhängigkeit / Professionalität	KNAW & NWO	IQSS (et al.)	Macmillan Publishers	ICPSR
Ausrichtung	Alle Disziplinen	Alle Disziplinen	Alle Disziplinen	Social political science
Jurisdiktion	Niederlande	USA	UK	USA
Aufnahmekriterien	Keine	Keine	Keine	social and political data
Datennutzung einschränkbar	✓	✓	✗	✓ bzw. ✗
Persistente Identifikatoren	✓	✓	✓	✓
Kostenfreie Datenübergabe	✓	✓	✓	✓
Kostenfreie Datennutzung	✓	✓	✓	✗ bzw. ✓

Technische Aspekte - Vergleich

	PsychData	GESIS datorium	Zenodo	OSF
Umwandlung von Formaten	✓	✗	✗	✗
Langfristige Speicherung	Mind. 8 Jahre, automatische Verlängerung (Lifetime)	Lifetime of repository (mind. 10 Jahre)	Lifetime of repository (mind. 20 Jahre)	Lifetime + 50 Jahre
Ablage zusätzlicher Materialien	✓	✓	✓	✓
Versionierung	✓	✗	✓	✓

Technische Aspekte - Vergleich

	DANS	Harvard Dataverse	figshare	openICPSR
Umwandlung von Formaten	✗	✓	✗	✗
Langfristige Speicherung	Longterm	Lifetime of IQSS	Lifetime of repository (mind. 10 Jahre)	OpenICPSR provides ... for the long term. ... further funding to extend preservation and public access. Thereafter, unfunded deposits transfer to the ICPSR membership. (https://www.openicpsr.org/openicpsr/fags)
Ablage zusätzlicher Materialien	✓	✓	✓	✓
Versionierung	✓	✓	✓	✓

Inhaltliche Aspekte - Vergleich

	PsychData	GESIS datorium	Zenodo	OSF
Studiendokumentation	Eigenes, an Psychologie angepasstes Schema	DDI	DC	✗
Standardisierte Variablendokumentation	✓	✗	✗	✗
Zitationspflicht/ Koautorenschaft	✓	✓	✓	✓
Beratung, Unterstützung, Anleitung	✓	✓	✗	✗

Inhaltliche Aspekte - Vergleich

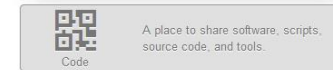
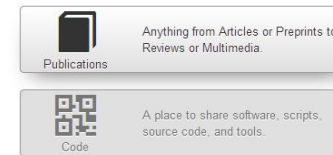
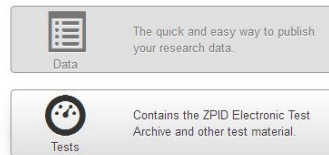
	DANS	Harvard Dataverse	figshare	openICPSR
Studiendokumentation	DDI	DDI	DC	DDI
Standardisierte Variablendokumentation	✗	✗	✗	✗
Zitationspflicht/ Koautorenschaft	✓	✓	(✓)	✓
Beratung, Unterstützung, Anleitung	✓	✓	✗	✓

Resümee: Unterschiede vor allem in folgenden Aspekten:

- Datennutzung ist einschränkbar
- Umwandlung von Dateiformaten
- Art (und Umfang) der Studiendokumentation
- Standardisierte Variablendokumentation
- Beratung, Unterstützung, Anleitung

- Minimum Kriterium: Peer-reviewed publication
- Kontaktaufnahme – psychdata@zpid.de
- Abschluss eines Datengebervertrags
- Übergabe von Studiendokumentation, Daten und Variablendokumentation (via myspsychdata.zpid.de)
- Überprüfung und Ergänzung der Materialien durch PsychData-Team
- Abschluss-Revision durch Datengeber
- Veröffentlichung und DOI-Vergabe
- Datenweitergabe nur für wissenschaftliche Zwecke (Zusatzvereinbarungen möglich)

Bald am ZPID: PsychArchives



- Repositorium für die Psychologie
 - Software
 - Forschungsdaten
 - Publikationen
 - Tests
- International
- Open access und kostenlos
- Digital Object Identifier (DOI)

14:00 – 14:45	Forschungsdatenmanagement und Bereitstellung von Daten
14:45 – 15:00	<i>Pause</i>
15:00 – 15:45	Datenaufbereitung
15:45 – 16:00	<i>Pause</i>
16:00 – 16:30	Einführung DataWiz
16:30 – 17:30	Hands-on DataWiz
17:30 – 18:00	Rückmeldung, Diskussion & Abschluss

- Forschungsdaten zur Nachprüfbarkeit sichern
 - Zusammenstellung des Archivierungspakets (Forschungsdaten, Dokumentation etc.)
 - Überführung von Daten in Archivierungsformate
 - (Finale) Aufbereitung der Daten
 - Nachnutzbarkeit der bereitgestellten Daten sicherstellen
 - Kriterienkatalog Anhang A: DGPs Empfehlungen
- Forschungsdaten für die Nachnutzung in einem Repositorium bereitstellen
 - Auswahl der Daten
 - DGPs Empfehlungen: Publikationstypen
 - Wahl eines geeigneten Repositorium
 - Datengebervertrag/Lizenzen festlegen
 - Übergabe der Daten



Anhang A

1. **Dauer der Aufbewahrung**
2. **Technische Zugänglichkeit**
3. **Dokumentation auf Variablenebene (Inhaltliche Verständlichkeit)**
4. **Nutzung persistenter Identifikatoren (Auffindbarkeit)**
5. **Dokumentation**
 1. auf **Studienebene**
 2. auf **Datenobjektebene**
6. **Bereitstellung von**
 1. **Instrumenten**
 2. **Software**
 3. **Materialien**

1. **Dauer der Aufbewahrung:** Die DFG schreibt vor, die Primärdaten *mindestens* 10 Jahre aufzubewahren. Diese Regelung leitet sich aus der Guten Wissenschaftlichen Praxis ab, die insbesondere auf die Vermeidung von wissenschaftlichem Fehlverhalten abzielt und sicherstellen soll, dass in einem Verdachtsfall die Forschungsergebnisse nachvollzogen und geprüft werden können. Das impliziert nicht, dass die Daten nach diesem Zeitraum gelöscht werden *sollten*. Da es, außer bei Studien mit extrem großen Datenmengen, kein Argument für die Löschung anonymisierter Daten gibt, befürwortet die DGPs im Normalfall eine unbegrenzte Sicherung ohne zeitgebundenen "Löschmechanismus" oder „Löschzwang“.
2. **Die Speicherung in einem nicht-proprietären Datenformat** (wie z.B. csv-Dateien oder Text-Dateien) ist einem proprietären Dateiformat (.mat-Datei, SPSS-Datei, SAS-Datei) vorzuziehen damit die Transparenz nicht auf Besitzer(innen) ggf. teurer Spezialsoftware beschränkt wird. Entsprechend sollen die Daten so aufbereitet sein, dass sie ohne Spezialsoftware auslesbar und nutzbar sind. Zum Beispiel würde ein binärer Output aus einer unbekannten Experimentalsoftware dem Ziel der Forschungstransparenz nicht genügen; hier sind die Daten soweit aufzubereiten, dass sie in einem nachvollziehbaren Datenformat vorliegen. Wenn das nicht möglich ist, sollte klar dokumentiert werden, mit welcher Software die Daten auslesbar sind.
3. **Neben der technischen Zugänglichkeit muss auch die inhaltliche Verständlichkeit der Daten gewährleistet sein.** Alle Variablen müssen in einem digitalen Codebuch dokumentiert sein²⁶. Es muss klar sein, welche manipulierte bzw. gemessene Variable in der Publikation zu welcher Variablen im Datensatz gehört. Idealerweise liegen den Daten auch Analyseskripte bei (z.B. R-Skripte oder SPSS-Syntax), die die publizierten Ergebnisse reproduzieren.
4. **Der Speicherort sollte einen persistenten Identifikator erhalten** (z.B. eine persistente URL oder falls möglich, einen DOI). Das erlaubt eine einheitliche Zitation der Daten. Der Speicherort soll in der Publikation, die auf diesen Daten beruht, genannt werden, so dass Forschungsdaten auch gefunden werden können.
5. **Neben der Dokumentation auf Variablenebene (Codebuch) ist auch eine Dokumentation auf Studienebene und auf Datenobjektebene** (Dateien, Versionen, etc.) erforderlich. Das kann zum Beispiel durch eine README-Datei im Repository gelöst werden, die einen Überblick über die archivierten Dateien gibt, und gegebenenfalls Hinweise zur Reproduktion gibt (z.B., Welche Software ist notwendig? In welcher Reihenfolge sind die Skripte auszuführen? Wie soll der Datensatz zitiert werden?).
6. **Es wird empfohlen, neben den Daten auch Instrumente, Software und Materialien**

Schönbrodt, Gollwitzer, & Abele-Brehm, 2016

Anhang A

1. **Dauer der Aufbewahrung**
2. **Technische Zugänglichkeit**
3. **Dokumentation auf Variablenebene (Inhaltliche Verständlichkeit)**
4. **Nutzung persistenter Identifikatoren (Auffindbarkeit)**
5. **Dokumentation**
 1. auf **Studienebene**
 2. auf **Datenobjektebene**
6. **Bereitstellung von**
 1. **Instrumenten**
 2. **Software**
 3. **Materialien**

1. Dauer der Aufbewahrung: Die DFG schreibt vor, die Primärdaten *mindestens* 10 Jahre aufzubewahren. Diese Regelung leitet sich aus der Guten Wissenschaftlichen Praxis ab, die insbesondere auf die Vermeidung von wissenschaftlichem Fehlverhalten abzielt und sicherstellen soll, dass in einem Verdachtsfall die Forschungsergebnisse nachvollzogen und geprüft werden können. Das impliziert nicht, dass die Daten nach diesem Zeitraum gelöscht werden *sollten*. Da es, außer bei Studien mit extrem großen Datenmengen, kein Argument für die Löschung anonymisierter Daten gibt, befürwortet die DGPs im Normalfall eine unbegrenzte Speicherung des Datensatzes als „Lebensdauerarchiv“ oder „Löschzwang“.
2. Die Speicherformate (Dateien oder Text-S-Datei) vorzuziehen, die nicht auf eine bestimmte Software beschränkt wird. Entsprechende Software sollte frei verfügbar sein und nutzbar sein. Experimentaldaten sollten so weit auf der Ebene der Datenobjektebene dokumentiert werden, wie es nicht möglich ist, wenn die Daten nicht möglich sind.
3. Neben der technischen Zugänglichkeit muss auch die inhaltliche Verständlichkeit der Daten gewährleistet sein. Alle Variablen müssen in einem digitalen Codebuch dokumentiert sein²⁶. Es muss klar sein, welche manipulierte bzw. gemessene Variable in der Publikation zu welcher Variablen im Datensatz gehört. Idealerweise liegen den Daten auch Analyseskripte bei (z.B. R-Skripte oder SPSS-Syntax), die die publizierten Ergebnisse reproduzieren.
4. Der Speicherort sollte einen persistenten Identifikator erhalten (z.B. eine persistente URL oder falls möglich, einen DOI). Das erlaubt eine einheitliche Zitation der Daten. Der Speicherort soll in der Publikation, die auf diesen Daten beruht, genannt werden, so dass Forschungsdaten auch gefunden werden können.
5. Neben der Dokumentation auf Variablenebene (Codebuch) ist auch eine Dokumentation auf Studienebene und auf Datenobjektebene (Dateien, Versionen, etc.) erforderlich. Das kann zum Beispiel durch eine README-Datei im Repository gelöst werden, die einen Überblick über die archivierten Dateien gibt, und gegebenenfalls Hinweise zur Reproduktion gibt (z.B., Welche Software ist notwendig? In welcher Reihenfolge sind die Skripte auszuführen? Wie soll der Datensatz zitiert werden?).
6. Es wird empfohlen, neben den Daten auch Instrumente, Software und Materialien

Wahl eines geeigneten
Repositoriums

Anhang A

1. **Dauer der Aufbewahrung**
2. **Technische Zugänglichkeit**
3. **Dokumentation auf Variablenebene (Inhaltliche Verständlichkeit)**
4. **Nutzung persistenter Identifikatoren (Auffindbarkeit)**
5. **Dokumentation**
 1. auf **Studienebene**
 2. auf **Datenobjektebene**
6. **Bereitstellung von**
 1. **Instrumenten**
 2. **Software**
 3. **Materialien**

1. Dauer der Aufbewahrung: Die DFG schreibt vor, die Primärdaten *mindestens* 10 Jahre aufzubewahren. Diese Regelung leitet sich aus der Guten Wissenschaftlichen Praxis ab, die insbesondere auf die Vermeidung von wissenschaftlichem Fehlverhalten abzielt und sicherstellen soll, dass in einem Verdachtsfall die Forschungsergebnisse nachvollzogen und geprüft werden können. Das impliziert nicht, dass die Daten nach diesem Zeitraum gelöscht werden *sollten*. Da es, außer bei Studien mit extrem großen Datenmengen, kein Argument für die Löschung anonymisierter Daten gibt, befürwortet die DGPs im Normalfall eine unbegrenzte Speicherung von Primärdaten²⁵.
Dateien oder Text- (z.B. CSV-Datei) vorzuziehen. Die Zugänglichkeit ist nicht beschränkt wird. Die Software auslesbar ist. Hier sind die Daten vorliegen. Wenn das die Daten auslesbar sind.
2. Die Speicherung von Dateien ist nicht möglich. Die Zugänglichkeit ist nicht beschränkt wird. Die Software auslesbar ist. Hier sind die Daten vorliegen. Wenn das die Daten auslesbar sind.
3. Neben der technischen Zugänglichkeit muss auch die inhaltliche Verständlichkeit der Daten gewährleistet sein. Alle Variablen müssen in einem digitalen Codebuch dokumentiert sein²⁶. Es muss klar sein, welche manipulierte bzw. gemessene Variable in der Publikation zu welcher Variablen im Datensatz gehört. Idealerweise liegen den Daten auch Analyseskripte bei (z.B. R-Skripte oder SPSS-Syntax), die die publizierten Ergebnisse reproduzieren.
4. Der Speicherort sollte einen persistenten Identifikator erhalten (z.B. eine persistente URL oder falls möglich, einen DOI). Das erlaubt eine einheitliche Zitation der Daten. Der Speicherort soll in der Publikation, die auf diesen Daten beruht, genannt werden, so dass Forschungsdaten auch gefunden werden können.
5. Neben der Dokumentation auf Variablenebene (Codebuch) ist auch eine Dokumentation auf Studienebene und auf Datenobjektebene (Dateien, Versionen, etc.) erforderlich. Das kann zum Beispiel durch eine README-Datei im Repository gelöst werden, die einen Überblick über die archivierten Dateien gibt, und gegebenenfalls Hinweise zur Reproduktion gibt (z.B., Welche Software ist notwendig? In welcher Reihenfolge sind die Skripte auszuführen? Wie soll der Datensatz zitiert werden?).
6. Es wird empfohlen, neben den Daten auch Instrumente, Software und Materialien

Kann nicht durch die Wahl
eines geeigneten
Repositoriums allein geleistet
werden.

Anhang A

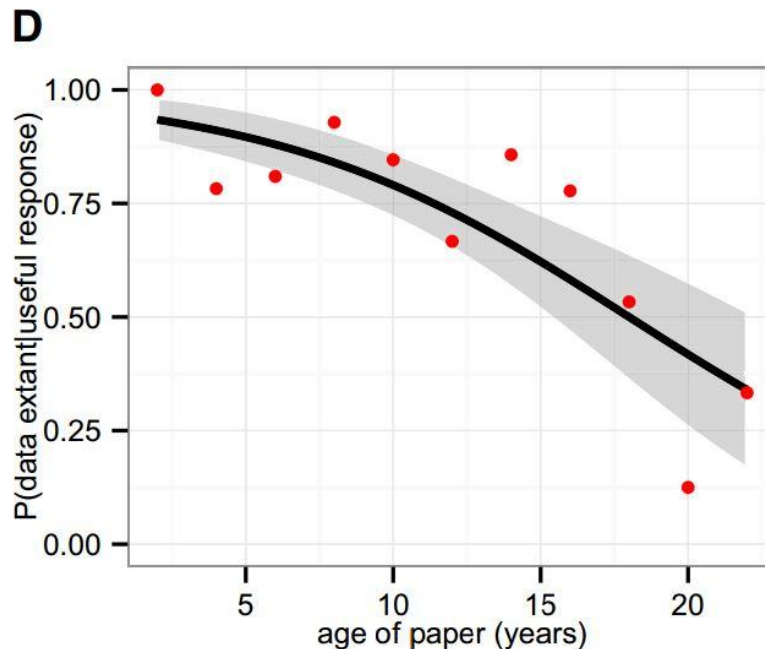
1. **Dauer der Aufbewahrung**
2. **Technische Zugänglichkeit**
3. **Dokumentation auf Variablenebene (Inhaltliche Verständlichkeit)**
4. **Nutzung persistenter Identifikatoren (Auffindbarkeit)**
5. **Dokumentation**
 1. auf **Studienebene**
 2. auf **Datenobjektebene**
6. **Bereitstellung von**
 1. **Instrumenten**
 2. **Software**
 3. **Materialien**

1. **Dauer der Aufbewahrung:** Die DFG schreibt vor, die Primärdaten *mindestens* 10 Jahre aufzubewahren. Diese Regelung leitet sich aus der Guten Wissenschaftlichen Praxis ab, die insbesondere auf die Vermeidung von wissenschaftlichem Fehlverhalten abzielt und sicherstellen soll, dass in einem Verdachtsfall die Forschungsergebnisse nachvollzogen und geprüft werden können. Das impliziert nicht, dass die Daten nach diesem Zeitraum gelöscht werden *sollten*. Da es, außer bei Studien mit extrem großen Datenmengen, kein Argument für die Löschung anonymisierter Daten gibt, befürwortet die DGPs im Normalfall eine unbegrenzte Sicherung ohne zeitgebundenen "Löschmechanismus" oder „Löschzwang“.
2. **Die Speicherung in einem nicht-proprietären Datenformat** (wie z.B. csv-Dateien oder Text-Dateien) ist einem proprietären Dateiformat (.mat-Datei, SPSS-Datei, SAS-Datei) vorzuziehen damit die Transparenz nicht auf Besitzer(innen) ggf. teurer Spezialsoftware beschränkt wird. Entsprechend sollen die Daten so aufbereitet sein, dass sie ohne Spezialsoftware auslesbar und nutzbar sind. Zum Beispiel würde ein binärer Output aus einer unbekannten Experimentalsoftware dem Ziel der Forschungstransparenz nicht genügen; hier sind die Daten soweit aufzubereiten, dass sie in einem nachvollziehbaren Datenformat vorliegen. Wenn das nicht möglich ist, sollte klar dokumentiert werden, mit welcher Software die Daten auslesbar sind.
3. **Neben der technischen Zugänglichkeit muss auch die inhaltliche Verständlichkeit der Daten gewährleistet sein.** Alle Variablen müssen in einem digitalen Codebuch dokumentiert sein²⁶. Es muss klar sein, welche manipulierte bzw. gemessene Variable in der Publikation zu welcher Variablen im Datensatz gehört. Idealerweise liegen den Daten auch Analyseskripte bei (z.B. R-Skripte oder SPSS-Syntax), die die publizierten Ergebnisse reproduzieren.
4. **Der Speicherort sollte einen persistenten Identifikator erhalten** (z.B. eine persistente URL oder falls möglich, einen DOI). Das erlaubt eine einheitliche Zitation der Daten. Der Speicherort soll in der Publikation, die auf diesen Daten beruht, genannt werden, so dass Forschungsdaten auch gefunden werden können.
5. **Neben der Dokumentation auf Variablenebene (Codebuch) ist auch eine Dokumentation auf Studienebene und auf Datenobjektebene** (Dateien, Versionen, etc.) erforderlich. Das kann zum Beispiel durch eine README-Datei im Repository gelöst werden, die einen Überblick über die archivierten Dateien gibt, und gegebenenfalls Hinweise zur Reproduktion gibt (z.B., Welche Software ist notwendig? In welcher Reihenfolge sind die Skripte auszuführen? Wie soll der Datensatz zitiert werden?).
6. **Es wird empfohlen, neben den Daten auch Instrumente, Software und Materialien**

Kann ich die Daten überhaupt selbst dauerhaft nutzen?

**Kann ich Anderen Daten so zur Verfügung stellen, dass sie
die Daten nutzen können und die Daten dauerhaft
zugänglich sind?**

“The underlying data researchers analyze to come to their published conclusions ... **becomes less and less accessible to researchers over the years.**” (Vines et al, 2014)



(D) Predicted probability that the data were extant (either “shared” or “exist but unwilling to share”) given that we received a useful response.

- Aktuelle SPSS Version: SPSS 15
 - SPSS Statistics (2008) 16, 17.0: major revision, user interface rewritten in Java (solved the problem of having different versions on different platform.
 - SPSS portable file format verliert an Relevanz
 - Einführung von spv-Output-Files (vorher spo)

SPSS Legacy Viewer

Product: SPSS Legacy Viewer

Date: Last updated: August 13, 2010

Title: LegacyViewer.zip

The SPSS Legacy Viewer allows IBM SPSS Statistics 16, 17, 18 and 19 for Windows users to view and edit output (.spo files) created in previous versions of SPSS for Windows. Please note that the Legacy Viewer can be installed on Windows Client operating systems only (XP, Vista and Windows 7)

URL: <ftp://ftp.software.ibm.com/software/analytics/spss/support/Stats/Docs/Statistics/LegacyViewer/readme.html>

**Proprietäre/herstellereigene
Binäre**

vs.
vs.

**nicht-proprietäre/offene Formate?
textbasierte Formate?**

The image shows four software editors side-by-side, illustrating different data formats. The top-left editor is 'ffcRData - Editor' showing a binary file. The top-right editor is 'txt.tab - Editor' showing a text-based table with columns 'FILE', 'FORMAT', 'COMMONS', and 'TAB'. The bottom-left editor is 'ffc.sav - Editor' showing a binary file. The bottom-right editor is 'ffc.xls - Editor' showing a spreadsheet. Arrows point from labels '.RData', '.sav', and '.xls' to a box containing '.txt | .tab | .csv | .xml'.

.RData

.sav

.xls

.txt | .tab | .csv | .xml

Anhang A

1. **Dauer der Aufbewahrung**
2. **Technische Zugänglichkeit**
3. **Dokumentation auf Variablenebene (Inhaltliche Verständlichkeit)**
4. **Nutzung persistenter Identifikatoren (Auffindbarkeit)**
5. **Dokumentation**
 1. auf **Studienebene**
 2. auf **Datenobjektebene**
6. **Bereitstellung von**
 1. **Instrumenten**
 2. **Software**
 3. **Materialien**

1. **Dauer der Aufbewahrung:** Die DFG schreibt vor, die Primärdaten *mindestens* 10 Jahre aufzubewahren. Diese Regelung leitet sich aus der Guten Wissenschaftlichen Praxis ab, die insbesondere auf die Vermeidung von wissenschaftlichem Fehlverhalten abzielt und sicherstellen soll, dass in einem Verdachtsfall die Forschungsergebnisse nachvollzogen und geprüft werden können. Das impliziert nicht, dass die Daten nach diesem Zeitraum gelöscht werden *sollten*. Da es, außer bei Studien mit extrem großen Datenmengen, kein Argument für die Löschung anonymisierter Daten gibt, befürwortet die DGPs im Normalfall eine unbegrenzte Sicherung ohne zeitgebundenen "Löschmechanismus" oder „Löschzwang“.
2. **Die Speicherung in einem nicht-proprietären Datenformat** (wie z.B. csv-Dateien oder Text-Dateien) ist einem proprietären Dateiformat (.mat-Datei, SPSS-Datei, SAS-Datei) vorzuziehen damit die Transparenz nicht auf Besitzer(innen) ggf. teurer Spezialsoftware beschränkt wird. Entsprechend sollen die Daten so aufbereitet sein, dass sie ohne Spezialsoftware auslesbar und nutzbar sind. Zum Beispiel würde ein binärer Output aus einer unbekannten Experimentalsoftware dem Ziel der Forschungstransparenz nicht genügen; hier sind die Daten soweit aufzubereiten, dass sie in einem nachvollziehbaren Datenformat vorliegen. Wenn das nicht möglich ist, sollte klar dokumentiert werden, mit welcher Software die Daten auslesbar sind.
3. **Neben der technischen Zugänglichkeit muss auch die inhaltliche Verständlichkeit der Daten gewährleistet sein.** Alle Variablen müssen in einem digitalen Codebuch dokumentiert sein²⁶. Es muss klar sein, welche manipulierte bzw. gemessene Variable in der Publikation zu welcher Variablen im Datensatz gehört. Idealerweise liegen den Daten auch Analyseskripte bei (z.B. R-Skripte oder SPSS-Syntax), die die publizierten Ergebnisse reproduzieren.
4. **Der Speicherort sollte einen persistenten Identifikator erhalten** (z.B. eine persistente URL oder falls möglich, einen DOI). Das erlaubt eine einheitliche Zitation der Daten. Der Speicherort soll in der Publikation, die auf diesen Daten beruht, genannt werden, so dass Forschungsdaten auch gefunden werden können.
5. **Neben der Dokumentation auf Variablenebene (Codebuch) ist auch eine Dokumentation auf Studienebene und auf Datenobjektebene** (Dateien, Versionen, etc.) erforderlich. Das kann zum Beispiel durch eine README-Datei im Repository gelöst werden, die einen Überblick über die archivierten Dateien gibt, und gegebenenfalls Hinweise zur Reproduktion gibt (z.B., Welche Software ist notwendig? In welcher Reihenfolge sind die Skripte auszuführen? Wie soll der Datensatz zitiert werden?).
6. **Es wird empfohlen, neben den Daten auch Instrumente, Software und Materialien**

1. Binäre Formate verbinden (häufig) Codebuch und Datenmatrix.
Textuelle Formate (csv-Dateien) können das (in der Regel) nicht.
2. Binäre Dateien können offen dokumentiert sein (z.B. PDF).
3. Drittprogramme für verbreitete proprietäre Formate (z.B. sav: PSPP, R-Pakete, DataWiz,...).

→ **theoretisch** ist mit proprietären inhaltliche Verständlichkeit besser sicherzustellen

	ID_rater	Sex_rater
1	S4	0
2	S5	0
3	S6	0
4	S7	0
5	S8	0
6	S9	0
7	S11	1
8	S12	1
9	S13	1
10	S14	1
11	S15	1
12	S16	1
13	S17	1
14	S18	1
15	S21	1
16	S22	1
17	S23	1
18	S24	1
19	S25	0
20	S1	0
21	S2	1
22	S3	0
23	S27	0
24	S28	0
25	S29	0
26	S31	0
27	S32	0
28	S33	1
29	S34	1
30	S35	0

1	ID_rater	Zeichenfolge	24	0		Keine	Keine	8
2	Sex_rater	Zeichenfolge	24	0		Keine	Keine	8
3	Age_rater	Zeichenfolge	24	0		Keine	Keine	8

Tan, D., Russell-Smith, S., Simons, J., Maybery, M., Leung, D., Ng, H., & Whitehouse, A. (2015). Perceived gender ratings for high and low scorers on the Autism-spectrum Quotient consistent with the extreme male brain account of autism [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.18599>

'Male faces data_final.sav' and 'Female faces data_final.sav' contain masculinity (for male faces) and femininity (for female faces) ratings of face photographs. Each photograph is labelled as a variable (top row, third cell onwards) and subsequent rows represent ratings provided by each rater (E.g., first row of data represents ratings provided by rater for each of the photograph presented to him/her). 'Male voices data_final.sav' and 'Female voices data_final.sav' contain masculinity (for male voices) and femininity (for female voices) ratings of voice recordings. Data layout is the same as face data described above. Click on variable tab for details on each variable. 'AQ_rating_corr_face.sav' and 'AQ_rating_corr_voice.sav' contain mean ratings of male and female faces and voices, and their respective AQ scores.

<https://doi.org/10.5281/zenodo.18599>

Wertbeschriftungen

Wert:

Rechtschreibung...

Beschriftung:

Hinzufügen

Ändern

Entfernen

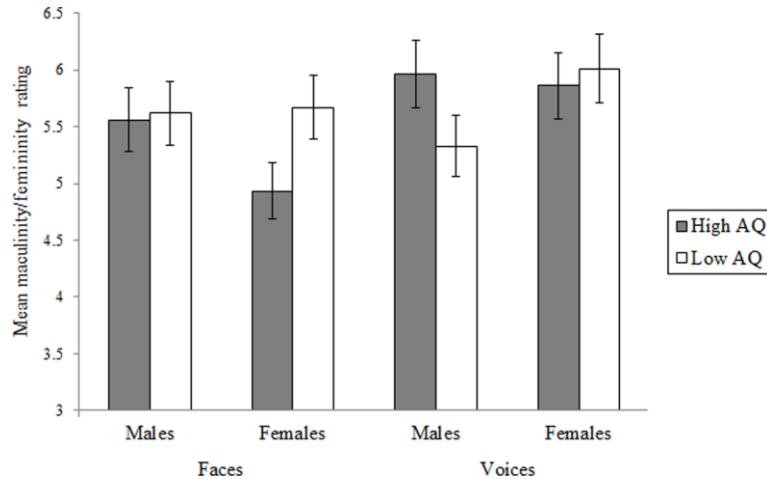
1,00 = "FHAQ"
2,00 = "FLAQ"
3,00 = "MHAQ"
4,00 = "MLAQ"

OK Abbrechen Hilfe

Tan, D., Russell-Smith, S., Simons, J., Maybery, M., Leung, D., Ng, H., & Whitehouse, A. (2015). Perceived gender ratings for high and low scorers on the Autism-spectrum Quotient consistent with the extreme male brain account of autism [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.18599>

'Male faces data_final.sav' and 'Female faces data_final.sav' contain masculinity (for male faces) and femininity (for female faces) ratings of face photographs. Each photograph is labelled as a variable (top row, third cell onwards) and subsequent rows represent ratings provided by each rater (E.g., first row of data represents ratings provided by rater for each of the photograph presented to him/her). 'Male voices data_final.sav' and 'Female voices data_final.sav' contain masculinity (for male voices) and femininity (for female voices) ratings of voice recordings. Data layout is the same as face data described above. Click on variable tab for details on each variable. 'AQ_rating_corr_face.sav' and 'AQ_rating_corr_voice.sav' contain mean ratings of male and female faces and voices, and their respective AQ scores.

<https://doi.org/10.5281/zenodo.18599>



OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Perceived Gender Ratings for High and Low Scorers on the Autism-Spectrum Quotient Consistent with the Extreme Male Brain Account of Autism

Diana Weiting Tan , Suzanna N. Russell-Smith , Jessica M. Simons, Murray T. Maybery, Doris Leung, Honey L. H. Ng, Andrew J. O. Whitehouse

Published: July 17, 2015 • <https://doi.org/10.1371/journal.pone.0131780>

Wertbeschriftungen

Wert:

Beschriftung:

- 1,00 = "FHAQ"
- 2,00 = "FLAQ"
- 3,00 = "MHAQ"
- 4,00 = "MLAQ"

1. Gültige Werte und ihre Kodierung sollte dokumentiert werden:
 - z.B. 0 = männlich, 1 = weiblich, 9 = nicht ausgefüllt
2. Für kontinuierliche Variablen sollte ein Bereich möglicher Werte definiert werden (falls möglich):
 - z.B. bei Leistungstests: 0 = Mindestpunktzahl, 30 = Maximalpunktzahl
 - Wenn nicht: z.B. Alter, Körpergröße
 - die Einheit sollte ersichtlich sein (Jahre, Meter, nmol/l).
3. Fehlende Werte sollten definiert werden (wenn sie kodiert wurden).
 - Bei mehr als einem fehlenden Wert, sollte ihre Kodierung dokumentiert werden
4. Zuordnung zu Fragebogenitems oder Ähnlichem sollte erkenntlich sein.
 1. Wenn Angabe des wörtlichen Text nicht möglich ist, sollte ein Verweis auf Itemnummer im Manual stattfinden: *Beck's Depression Inventory: [Manual Item number]*
 2. Berechnung von abgeleiteten Variablen sollte dokumentiert werden, ggf. auch auf Auswertungsvorschrift im Manual hingewiesen werden.

Anhang A

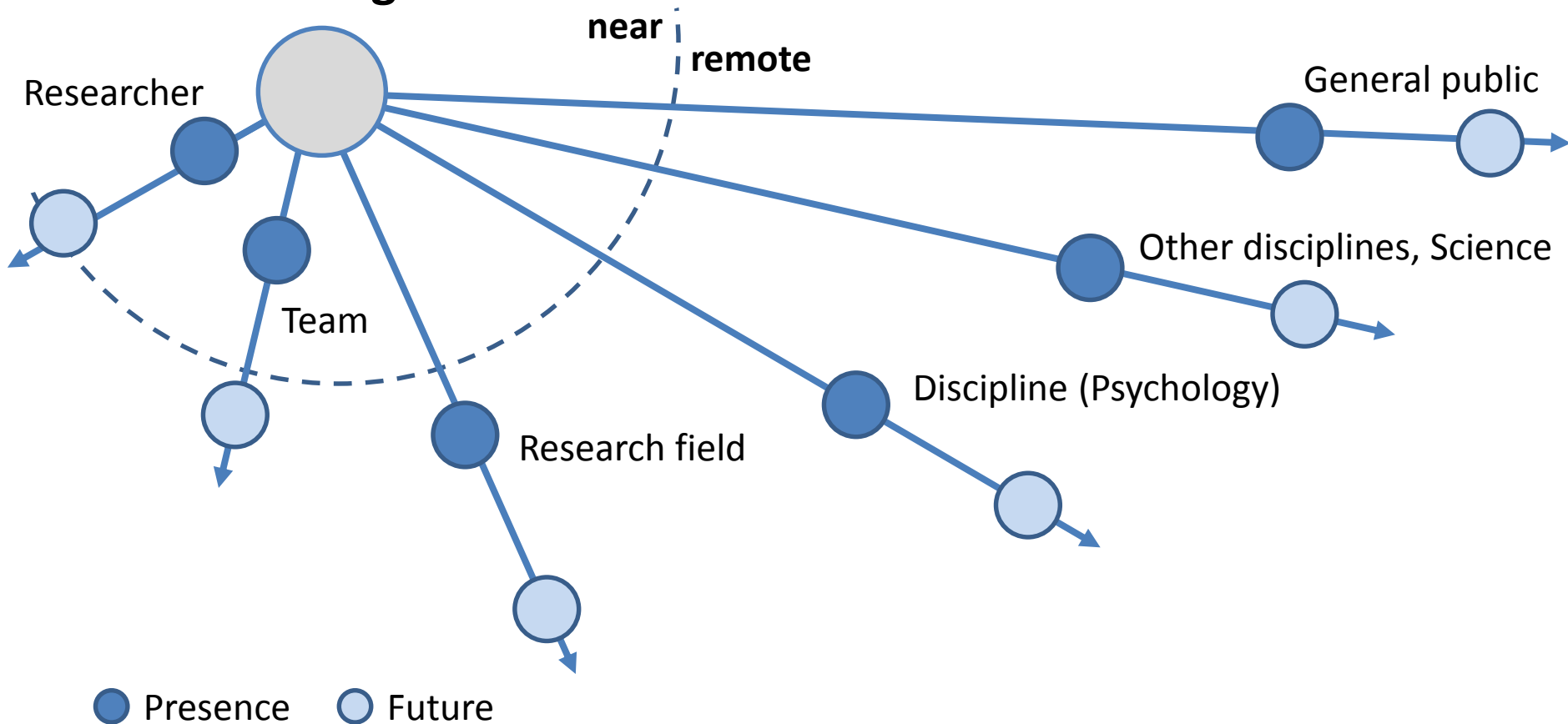
1. Dauer der Aufbewahrung
2. Technische Zugänglichkeit
3. Dokumentation auf Variablenebene (Inhaltliche Verständlichkeit)
4. Nutzung persistenter Identifikatoren (Auffindbarkeit)
5. Dokumentation
 1. auf Studienebene
 2. auf Datenobjektebene
6. Bereitstellung von
 1. Instrumenten
 2. Software
 3. Materialien

1. Dauer der Aufbewahrung: Die DFG schreibt vor, die Primärdaten *mindestens* 10 Jahre aufzubewahren. Diese Regelung leitet sich aus der Guten Wissenschaftlichen Praxis ab, die insbesondere auf die Vermeidung von wissenschaftlichem Fehlverhalten abzielt und sicherstellen soll, dass in einem Verdachtsfall die Forschungsergebnisse nachvollzogen und geprüft werden können. Das impliziert nicht, dass die Daten nach diesem Zeitraum gelöscht werden *sollten*. Da es, außer bei Studien mit extrem großen Datenmengen, kein Argument für die Löschung anonymisierter Daten gibt, befürwortet die DGPs im Normalfall eine unbegrenzte Sicherung ohne zeitgebundenen "Löschmechanismus" oder „Löschzwang“.
2. Die Speicherung in einem nicht-proprietären Datenformat (wie z.B. csv-Dateien oder Text-Dateien) ist einem proprietären Dateiformat (.mat-Datei, SPSS-Datei, SAS-Datei) vorzuziehen damit die Transparenz nicht auf Besitzer(innen) ggf. teurer Spezialsoftware beschränkt wird. Entsprechend sollen die Daten so aufbereitet sein, dass sie ohne Spezialsoftware auslesbar und nutzbar sind. Zum Beispiel würde ein binärer Output aus einer unbekannten Experimentalsoftware dem Ziel der Forschungstransparenz nicht genügen; hier sind die Daten soweit aufzubereiten, dass sie in einem nachvollziehbaren Datenformat vorliegen. Wenn das nicht möglich ist, sollte klar dokumentiert werden, mit welcher Software die Daten auslesbar sind.
3. Neben der technischen Zugänglichkeit muss auch die inhaltliche Verständlichkeit der Daten gewährleistet sein. Alle Variablen müssen in einem digitalen Codebuch dokumentiert sein²⁶. Es muss klar sein, welche manipulierte bzw. gemessene Variable in der Publikation zu welcher Variablen im Datensatz gehört. Idealerweise liegen den Daten auch Analyseskripte bei (z.B. R-Skripte oder SPSS-Syntax), die die publizierten Ergebnisse reproduzieren.
4. Der Speicherort sollte einen persistenten Identifikator erhalten (z.B. eine persistente URL oder falls möglich, einen DOI). Das erlaubt eine einheitliche Zitation der Daten. Der Speicherort soll in der Publikation, die auf diesen Daten beruht, genannt werden, so dass Forschungsdaten auch gefunden werden können.
5. Neben der Dokumentation auf Variablenebene (Codebuch) ist auch eine Dokumentation auf Studienebene und auf Datenobjektebene (Dateien, Versionen, etc.) erforderlich. Das kann zum Beispiel durch eine README-Datei im Repository gelöst werden, die einen Überblick über die archivierten Dateien gibt, und gegebenenfalls Hinweise zur Reproduktion gibt (z.B., Welche Software ist notwendig? In welcher Reihenfolge sind die Skripte auszuführen? Wie soll der Datensatz zitiert werden?).
6. Es wird empfohlen, neben den Daten auch Instrumente, Software und Materialien

Was brauchen wir, um Daten dauerhaft verständlich zu machen?


Welchen Grad an Studiendokumentation sollte mein Repository unterstützen?

Data Origin



Distance from data-origin as a function of target group and time

- **Was soll meine Dokumentation leisten? Reproduktion oder Beantwortung einer neuen Fragestellung durch Sekundäranalyse?**
- Ist ein Artikel alleine ausreichend?
 - Falls ein Artikel zum Verständnis erforderlich ist: Ist der Artikel frei verfügbar? (vgl. technische Zugänglichkeit)
 - Dokumentationsstandards beachten, sofern es sie für mein Teilgebiet gibt (z.B. [biosharing.org](https://www.biosharing.org/), [MRI-Daten](#), [Omics-Daten](#)).
 - In Ermangelung von Data Standards muss man hierbei (häufig) auf Reporting Standards für Printpublikationen zurückgreifen
- **Rechtliche Aspekte bei Bereitstellung von Fragebögen oder Artikeln beachten**

- Kontrollierte Vokabulare
- Nutzung von  Metadatenstandards
- Data Documentation Initiative (DDI)
 - The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences.
 - [DDI 3.2](#) oder DDI 2.5

Dublin Core Export

```
<?xml version='1.0' encoding='utf-8'?>
<dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xm
lns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:creator>Tan, Diana</dc:creator>
  <dc:creator>Russell-Smith, Suzanna</dc:creator>
  <dc:creator>Simons, Jessica</dc:creator>
  <dc:creator>Mayberry, Murray</dc:creator>
  <dc:creator>Leung, Doris</dc:creator>
  <dc:creator>Ng, Honey</dc:creator>
  <dc:creator>Whitehouse, Andrew</dc:creator>
  <dc:date>2015-06-16</dc:date>
  <dc:description>&lt;p&gt;Software required: SPSS&lt;/p&gt;

&lt;p&gt;&lt;p&gt;Male faces data_final.sav&lt;/p&gt; and &lt;p&gt;Female faces data_final.sav&lt;/p&gt; contain mascu
linity (for male faces) and femininity (for female faces) ratings of face photographs. Each photograph is labelled
as a variable (top row, third cell onwards) and subsequent rows represent ratings provided by each rater (E.g., first
row of data represents ratings provided by rater for each of the photograph presented to him/her). &lt;p&gt;Male voi
ces data_final.sav&lt;/p&gt; and &lt;p&gt;Female voices data_final.sav&lt;/p&gt; contain masculinity (for male voices)
and femininity (for female voices) ratings of voice recordings. Data layout is the same as face data described above
. Click on variable tab for details on each variable. &lt;p&gt;AQ_rating_corr_face.sav&lt;/p&gt; and &lt;p&gt;AQ_rati
ng_corr_voice.sav&lt;/p&gt; contain mean ratings of male and female faces and voices, and their respective AQ scores.
&lt;/p&gt;</dc:description>
  <dc:identifier>10.5281/zenodo.18599</dc:identifier>
  <dc:identifier>https://zenodo.org/record/18599</dc:identifier>
  <dc:identifier>oai:zenodo.org:18599</dc:identifier>
  <dc:rights>info:eu-repo/semantics/openAccess</dc:rights>
  <dc:rights>https://creativecommons.org/publicdomain/zero/1.0/</dc:rights>
  <dc:subject>autism</dc:subject>
  <dc:subject>autism spectrum disorder</dc:subject>
  <dc:subject>extreme male brain</dc:subject>
  <dc:subject>androgyny</dc:subject>
  <dc:subject>masculinity</dc:subject>
  <dc:subject>femininity</dc:subject>
  <dc:title>Perceived gender ratings for high and low scorers on the Autism-spectrum Quotient consistent with the ex
treme male brain account of autism</dc:title>
  <dc:type>info:eu-repo/semantics/other</dc:type>
</oai_dc:dc>
```

Export

[BibTeX](#) [CSL](#) [DataCite](#) [Dublin Core](#) [JSON](#)
[MARCXML](#) [Mendeley](#)

- Vorteile
 - Standardisiert
 - Maschinenlesbar
 - Interoperabel
- Nachteile
 - Extrem aufwändige Einarbeitung
- Lösung
 - Tools nutzen, die die angeleitete Dokumentation der Daten feldbasiert zulassen und Export nach DDI ermöglichen
 - z.B. DataWiz oder Nesstar Publisher
 - Daten und Metadaten standardisiert exportieren und in Archiv hochladen

Dublin Core Export

```
<?xml version="1.0" encoding="utf-8"?>
<dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:creator>Tan, Diana</dc:creator>
  <dc:creator>Russell-Smith, Suzanna</dc:creator>
  <dc:creator>Simons, Jessica</dc:creator>
  <dc:creator>Haybery, Murray</dc:creator>
  <dc:creator>Leung, Doris</dc:creator>
  <dc:creator>Ng, Honey</dc:creator>
  <dc:creator>Whitehouse, Andrew</dc:creator>
  <dc:date>2015-06-16</dc:date>
  <dc:description><p><Software required: SPSS</p>

  <p><Male faces data_final.sav<#> and <Female faces data_final.sav<#> contain masculinity (for male faces) and femininity (for female faces) ratings of face photographs. Each photograph is labelled as a variable (top row, third cell onwards) and subsequent rows represent ratings provided by each rater (e.g., first row of data represents ratings provided by rater for each of the photograph presented to him/her). <Male voices data_final.sav<#> and <Female voices data_final.sav<#> contain masculinity (for male voices) and femininity (for female voices) ratings of voice recordings. Data layout is the same as face data described above. Click on variable tab for details on each variable. <AQ_rating_corr_face.sav<#> and <AQ_rating_corr_voice.sav<#> contain mean ratings of male and female faces and voices, and their respective AQ scores.
  </p></dc:description>
  <dc:identifier>10.5281/zenodo.18599</dc:identifier>
  <dc:identifier>https://zenodo.org/record/18599</dc:identifier>
  <dc:identifier>oai:zenodo.org:18599</dc:identifier>
  <dc:rights>info:eu-repo/semantics/openaccess</dc:rights>
  <dc:rights>https://creativecommons.org/publicdomain/zero/1.0/</dc:rights>
  <dc:subject>autism</dc:subject>
  <dc:subject>autism spectrum disorder</dc:subject>
  <dc:subject>extreme male brain</dc:subject>
  <dc:subject>androgyny</dc:subject>
  <dc:subject>masculinity</dc:subject>
  <dc:subject>femininity</dc:subject>
  <dc:title>Perceived gender ratings for high and low scorers on the Autism-spectrum Quotient consistent with the extreme male brain account of autism</dc:title>
  <dc:type>info:eu-repo/semantics/other</dc:type>
</dc:dc>
```

Export

[BibTeX](#) [CSL](#) [DataCite](#) [Dublin Core](#) [JSON](#)
[MARCXML](#) [Mendeley](#)

Anhang A

1. Dauer der Aufbewahrung
2. Technische Zugänglichkeit
3. Dokumentation auf Variablenebene (Inhaltliche Verständlichkeit)
4. Nutzung persistenter Identifikatoren (Auffindbarkeit)
5. Dokumentation
 1. auf Studienebene
 2. auf Datenobjektebene
6. Bereitstellung von
 1. Instrumenten
 2. Software
 3. Materialien

1. Dauer der Aufbewahrung: Die DFG schreibt vor, die Primärdaten *mindestens* 10 Jahre aufzubewahren. Diese Regelung leitet sich aus der Guten Wissenschaftlichen Praxis ab, die insbesondere auf die Vermeidung von wissenschaftlichem Fehlverhalten abzielt und sicherstellen soll, dass in einem Verdachtsfall die Forschungsergebnisse nachvollzogen und geprüft werden können. Das impliziert nicht, dass die Daten nach diesem Zeitraum gelöscht werden *sollten*. Da es, außer bei Studien mit extrem großen Datenmengen, kein Argument für die Löschung anonymisierter Daten gibt, befürwortet die DGPs im Normalfall eine unbegrenzte Sicherung ohne zeitgebundenen "Löschmechanismus" oder „Löschzwang“.
2. Die Speicherung in einem nicht-proprietären Datenformat (wie z.B. csv-Dateien oder Text-Dateien) ist einem proprietären Dateiformat (.mat-Datei, SPSS-Datei, SAS-Datei) vorzuziehen damit die Transparenz nicht auf Besitzer(innen) ggf. teurer Spezialsoftware beschränkt wird. Entsprechend sollen die Daten so aufbereitet sein, dass sie ohne Spezialsoftware auslesbar und nutzbar sind. Zum Beispiel würde ein binärer Output aus einer unbekannten Experimentalsoftware dem Ziel der Forschungstransparenz nicht genügen; hier sind die Daten soweit aufzubereiten, dass sie in einem nachvollziehbaren Datenformat vorliegen. Wenn das nicht möglich ist, sollte klar dokumentiert werden, mit welcher Software die Daten auslesbar sind.
3. Neben der technischen Zugänglichkeit muss auch die inhaltliche Verständlichkeit der Daten gewährleistet sein. Alle Variablen müssen in einem digitalen Codebuch dokumentiert sein²⁶. Es muss klar sein, welche manipulierte bzw. gemessene Variable in der Publikation zu welcher Variablen im Datensatz gehört. Idealerweise liegen den Daten auch Analyseskripte bei (z.B. R-Skripte oder SPSS-Syntax), die die publizierten Ergebnisse reproduzieren.
4. Der Speicherort sollte einen persistenten Identifikator erhalten (z.B. eine persistente URL oder falls möglich, einen DOI). Das erlaubt eine einheitliche Zitation der Daten. Der Speicherort soll in der Publikation, die auf diesen Daten beruht, genannt werden, so dass Forschungsdaten auch gefunden werden können.
5. Neben der Dokumentation auf Variablenebene (Codebuch) ist auch eine Dokumentation auf Studienebene und auf Datenobjektebene (Dateien, Versionen, etc.) erforderlich. Das kann zum Beispiel durch eine README-Datei im Repository gelöst werden, die einen Überblick über die archivierten Dateien gibt, und gegebenenfalls Hinweise zur Reproduktion gibt (z.B., Welche Software ist notwendig? In welcher Reihenfolge sind die Skripte auszuführen? Wie soll der Datensatz zitiert werden?).
6. Es wird empfohlen, neben den Daten auch Instrumente, Software und Materialien



URL: <http://petapixel.com/2015/07/22/this-is-how-everyone-names-their-final-psd-files/>

- Interne **Versionierung und Namenskonventionen** müssen frühzeitig mitgeplant werden. -> Datenmanagementplanung
- Tools zur automatischen Versionierung nutzen (GitHub, OSF, DataWiz).
- **Versionierungskonzept ggf. bei der Wahl des Repositoriums berücksichtigen**

Inhaltliche
Verfü
siche

Dokumentation
der Daten gegen

Technische
Verfügbarkeit
llen

Dokum
star
ke


gen an
gement
ennen

Versi

he
ngen
n





- **Ziel:**
 - Entwicklung eines Assistenzsystems zur forschungsbegleitenden Unterstützung psychologischer Forscher in ihrem Datenmanagement
- **Förderung:** 
- **Laufzeit:** 01.11-2015 – 31.10.2017
- **Projektleiter:** PD Dr. Erich Weichselgartner

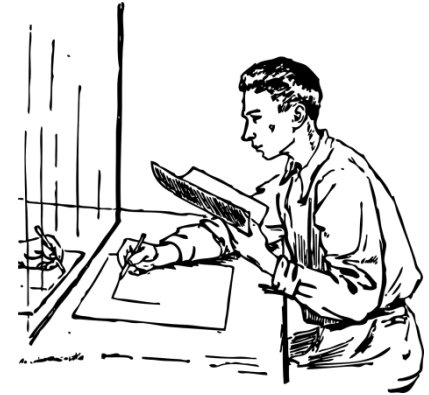
14:00 – 14:45	Forschungsdatenmanagement und Bereitstellung von Daten
14:45 – 15:00	<i>Pause</i>
15:00 – 15:45	Datenaufbereitung
15:45 – 16:00	<i>Pause</i>
16:00 – 16:30	Einführung DataWiz
16:30 – 17:30	Hands-on DataWiz
17:30 – 18:00	Rückmeldung, Diskussion & Abschluss

DataWiz

Assistenzsystems zum Management psychologischer
Forschungsdaten

- **Anforderungen**

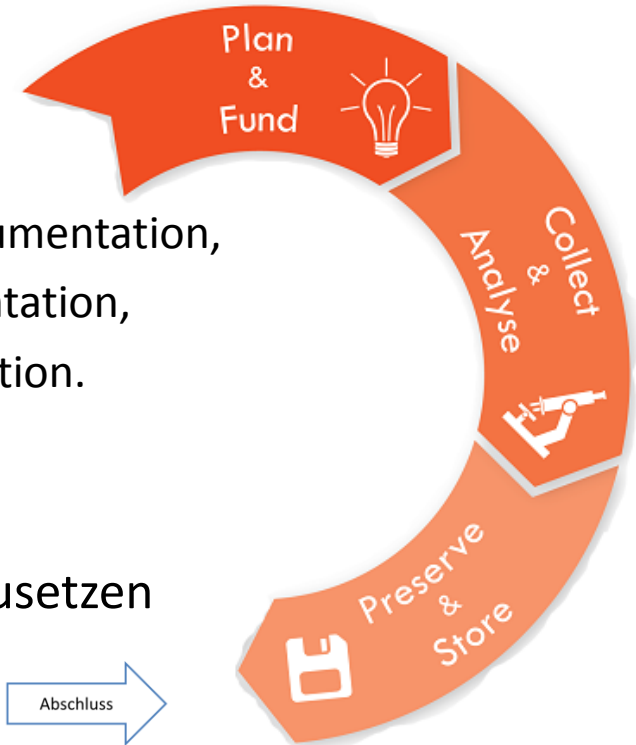
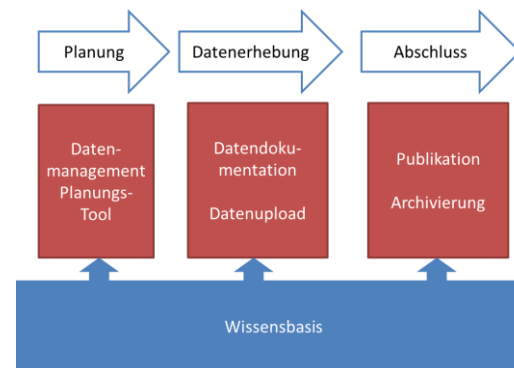
- Verfügbarkeit und **Interpretierbarkeit** der Daten stärken
- **Fachspezifische** Erfordernisse berücksichtigen
- **Einfacher Zugriff** auf **Informationen** zum Datenmanagement (wissensbasierte Unterstützung)
- Unterstützung bei der **Bewältigung zentraler Aufgaben** des Datenmanagements (prozedurale Unterstützung)
- Unterstützung frühzeitig und über den **gesamten Forschungsprozess** hinweg
- **Kontrolle** über die Daten beim Forschenden

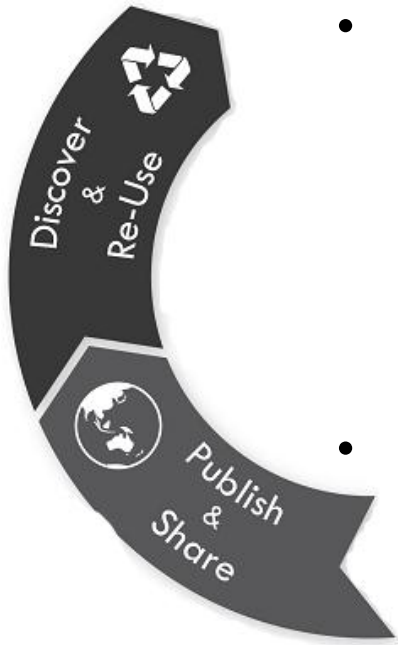


Insgesamt: Aufwand für adäquates Forschungsdatenmanagement minimieren

Was DataWiz ist...

- DataWiz verbindet
 - (1) Datenobjektdokumentation mit Variablendokumentation,
 - (2) Variablendokumentation mit Studiendokumentation,
 - (3) Studiendokumentation mit Projektdokumentation.
- Kollaboratives Arbeiten erleichtern
- Möglichkeit im eigenen Institut eine Instanz aufzusetzen
- Vermeidung von Redundanz in...
 - Datenmanagementplanung
 - Datendokumentation
 - Datenpublikation





- DataWiz ist **kein** öffentliches Repository. Es bereitet Daten für ein Repository vor.
 - DataWiz kann damit **nicht** (alleine) die langfristige technische Verfügbarkeit und Auffindbarkeit der Daten sicherstellen.
 - DataWiz vergibt keine PIDs.
- Workflow
 - Daten mit DataWiz dokumentieren und aufbereiten
 - Daten exportieren und in ein beliebiges Repository einstellen
 - Technische Verfügbarkeit und Auffindbarkeit als Aufgabe des Archivs

The image shows two overlapping elements. On the left is the cover of the 'FAKT-II' manual by Helfried Moosbrugger and Frank Goldhammer, published by Huber. The title is 'Frankfurter Adaptiver Konzentrationsleistungs-Test II'. Below the title, there is a section 'Geistige Eigentumsrechte Dritter:' with a 'Nein' button and an information icon. On the right is a screenshot of the DataWiz website. The website has a blue header with the DataWiz logo and the tagline 'manage your data'. Below the header is a navigation bar with links: Home, Before Collection, Data Collection, and After Collection. The main content area is titled 'Copyright of Third Parties' and contains a detailed explanation of copyright considerations for researchers, including a list of bullet points. A sidebar on the left of the website lists various topics like Data Management Plans, Reusing Datasets, and Privacy. At the bottom of the website screenshot is the Psychology Information ZPID logo.

Einsatzbereich

Der FAKT-II dient der adaptiven Ermittlung der individuellen Konzentrationsfähigkeit. Die Anwendungsbereiche des FAKT-II liegen in der Eignungsdiagnostik, der Arbeits- und Betriebspsychologie, Klinischen Psychologie, Schulpsychologie, Pädagogischen Psychologie, Verkehrspsychologie, Sportpsychologie, im neuro- und pharmakopsychologischen Bereich, in der Psychiatrie sowie in der Grundlagenforschung.

My Project 2

Beschreibung:

noch keine Projektbeschreibung vorhanden

Forscher:

noch keine Forscher eingetragen

Studienübersicht:

Studien anzeigen/ausblenden

My Study

Beschreibung:

noch keine Studienbeschreibung vorhanden

Forscher:

noch keine Forscher eingetragen

My Study 2

Beschreibung:

noch keine Studienbeschreibung vorhanden

Forscher:

noch keine Forscher eingetragen

Projektmitarbeiter:

Projektmitarbeiter anzeigen/ausblenden

Max Mustermann(datawiz@zpid.de)	Projektadministrator
Martin Kerwer(testuser@zpid.de)	Projektadministrator *
Sarah Stroh(stroh@zpid.de)	Projektadministrator

letztes Update am: 27.04.2017 13:52 durch Max Mustermann

Codebook

Datenmatrix

Name	Typ	Beschriftung	Werte	Fehlend
<input type="text" value="rnorm1990"/>	Numerisch	<input type="text" value="Index"/>	1 = "oauÖAÖS" 2 = "hallo"	98 - 99, 97
<input type="text" value="rnorm.1999..1"/>	Numerisch	<input type="text" value="Index"/>		1, 2
<input type="text" value="rnorm.1999"/>	Numerisch	<input type="text" value="Indexwert"/>	1.07804101285661 = "test" 34.0 = "tdg"	1.0

rnorm1990	rnorm.1999..1	rnorm.1999	test2
	-1		34.0
2	-1	-0.440175685888252	34.0
3	1	-0.109139785366486	5.0
4	1	0.778750192944401	
5	0	-0.266895826594832	6.0
6	-1	0.148958876555503	
7	-1	-0.315678964247938	6.0
8	-1	0.530066902222633	
9	-1	1.07584288909655	
10	-1	0.116160804667026	123213.23

Uploadbereiche auf Projekt und Studienebene

Dateien hochladen

Sichern Sie wichtige Dateien und teilen Sie sie mit Ihren Kollegen.
Es gibt **getrennte Upload-Bereiche** für Materialien auf **Projektebene** (z.B. Projektanträge, Projektberichte, Förderaufträge, Hintergrundliteratur) und **Studienebene** (z.B. Vorlagen für Einwilligungserklärungen oder Rohdaten einer Erhebung).
Datenmatrizen (z.B. .csv oder .sav) sollten auf Studienebene unter **Datensätze** hochgeladen und dokumentiert werden.
In der **Projekt- und Studiendokumentation** können Sie auf die hochgeladenen Dateien **verweisen**.
[Zum User Guide](#)

Schieben Sie Ihre Dateien hierhin,
oder klicken Sie auf das Feld.
Anschließend klicken Sie auf **Datei(en) hochladen**, um den Upload abzuschließen.
Die maximale Dateigröße beträgt **1,5 Gigabyte**.

Uploadfeld leeren

Datei(en) hochladen

Bit-Stream Preservation
beliebiger digitaler Inhalte
(z.B. Videoaufnahmen, EEG-
Daten, fMRT-Daten,
Screenshots)

14:00 – 14:45	Forschungsdatenmanagement und Bereitstellung von Daten
14:45 – 15:00	<i>Pause</i>
15:00 – 15:45	Datenaufbereitung
15:45 – 16:00	<i>Pause</i>
16:00 – 16:30	Einführung DataWiz
16:30 – 17:30	Hands-on DataWiz
17:30 – 18:00	Rückmeldung, Diskussion & Abschluss

Vielen Dank für Ihre Aufmerksamkeit!



Kontakt

DataWiz – DFG-Projekt am Forschungsdatenzentrum für die Psychologie (PsychData) des ZPID

- Erich Weichselgartner, Ltg. (weichselgartner@zpid.de)
- Ronny Bölder (boelder@zpid.de)
- Ina Dehnhard (dehnhard@zpid.de)
- Armin Günther (armin.guenther@zpid.de)
- Martin Kerwer (kerwer@zpid.de)

Schönbrodt, F., Gollwitzer, M., und Abele-Brehm, A. (2016). Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien. Abgerufen von

https://www.dgps.de/fileadmin/documents/Empfehlungen/Richtlinien_zum_Umgang_mit_Forschungsdaten_20160929.pdf

Tan, D., Russell-Smith, S., Simons, J., Maybery, M., Leung, D., Ng, H., & Whitehouse, A. (2015). Perceived gender ratings for high and low scorers on the Autism-spectrum Quotient consistent with the extreme male brain account of autism [Data set]. Zenodo.

<https://doi.org/10.5281/zenodo.18599>

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., ... Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology* 24(1): 94–97.