



# Forschungsdatenmanagement in der empirischen Bildungsforschung

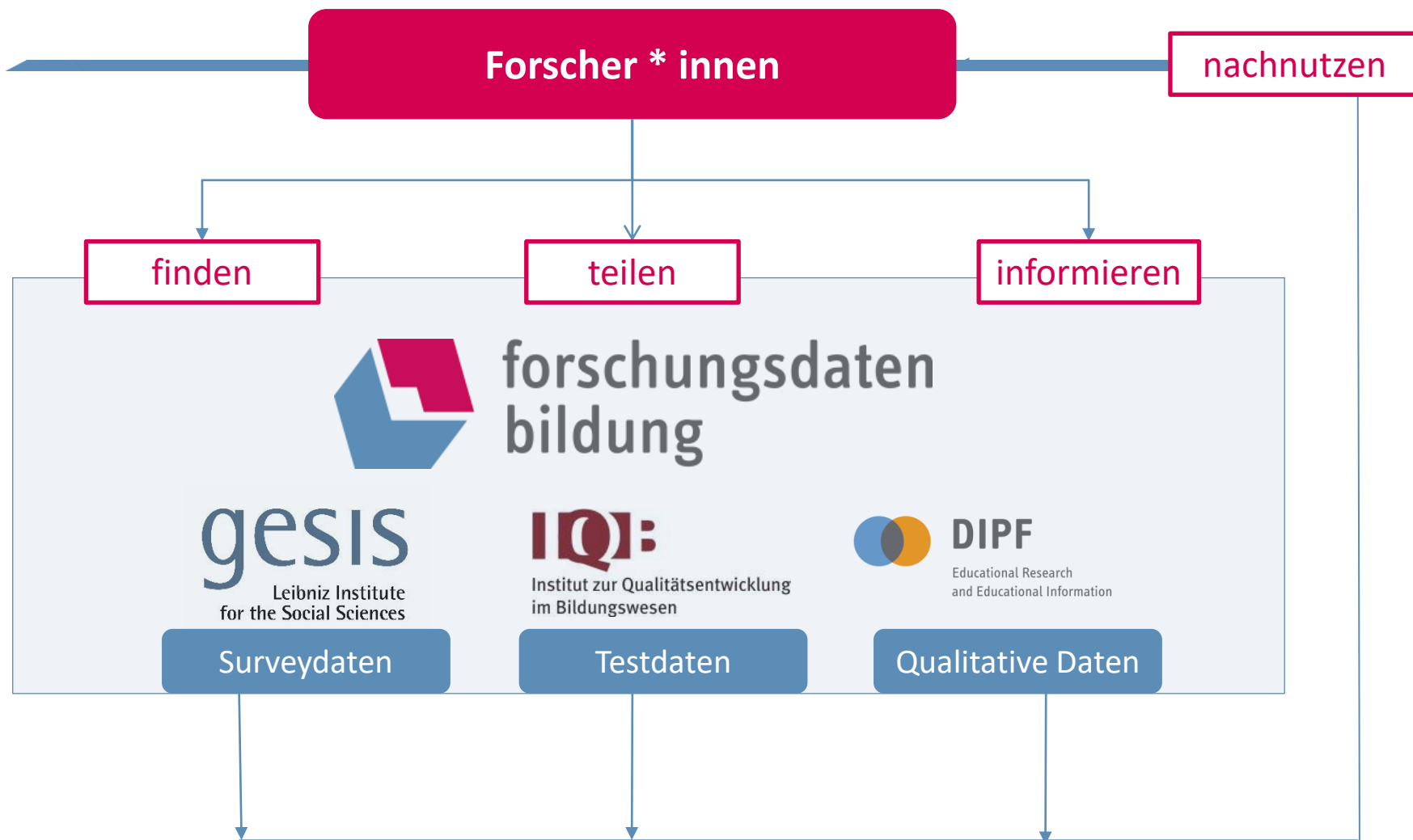
*Dr. Marcus Eisentraut*

GESIS - Leibniz-Institut für Sozialwissenschaften

Workshopreihe der BMBF-Förderline  
„INVITE“  
14.12.2022



# Der Verbund und seine Services



finden

**forschungsdaten bildung**  
DIFF  
Suchen & Finden | DATEN TEILEN | TIPPS & MATERIALIEN | ÜBER DEN VERBUND

**STUDIEN**  
forschungsdaten bildung.de gibt einen zentralen Überblick zu Forschungsprojekten aus dem Bereich der empirischen Bildungsforschung. Angebote werden jeweils allgemeine Studien- und Projektinformationen, Angaben zu verfügbaren Forschungsdaten, angelegten Erhebungs- und Auswertungsinstrumenten sowie Publikationen. Sie haben die Möglichkeit die hier verfügbare Gesamtheit aller nachgewiesenen Studien zu themen bzw. geistes bzw. gesellschaftswissenschaftlichen Disziplinen zu durchsuchen. Stellen Forschungsdaten zum Zweck der Nachnutzung zur Verfügung, werden Sie zum entsprechenden Anbieter weitergeleitet. Darunter werden die Datenzentren von GEOS, das FÖZ am IQB und das FÖZ Bildung am DIFF berücksichtigt. Angaben zu Zugangs- und Nutzungsbedingungen können Sie direkt beim jeweiligen Anbieter einsehen.

**STUDIENSUCHE**  
alle Wörter

10 Treffer pro Seite

**ÄNDERN PHÄNOMENE DAS INTERESSE?**  
Analyse der Konnotationen der Physik bei Schülerinnen und Schülern als Folge phänomenbasierten Unterrichtes  
Veröffentlicht: 01.02.2019 - 10.11.2019  
Von einem phänomenbasierten Physikunterricht, der von den Wahrnehmungen und alltäglichen Erfahrungen ausgeht und darauf abzielt, die physikalische Begriffswelt erschaffen, wird ein positiver Einfluss auf das Interesse von Schülern unterstellt. Für diese Normalität beg... > mehr

**ANSATZ FÜR EIN KLEINRÄUMIGES BILDUNGSMONITORING**  
Eine Analyse des Zusammenhangs zwischen sozialökonomischer Differenzierung und Bildungsnachvollziehung  
Veröffentlicht: 01.04.2008 - 31.08.2011  
Auf Basis von kleinräumiger Kommunalstatistik und der amtlichen Schulstatistik wird im Rahmen dieses Projekts der Zusammenhang zwischen sozialökonomischer Differenzierung und Bildungsnachvollziehung am Beispiel von 14 Städten des Ruhrgebiets analysiert und die Ber... > mehr

**FILTERN NACH:**  
Erhebungsmethoden  
 Messungen und Tests (67)  
 Eigenständig Ausgeführter Fragebogen (60)  
 Interview (23)  
 Beobachtung (21)  
 Fragebogen (7)  
> mehr...  
Analyseebenen  
 Schüler (67)  
 Lehrkräfte (44)  
 Eltern (24)  
 Auszubildende (20)  
 Schulleitung (20)  
> mehr...  
Schlagwörter  
 Deutschland (190)  
 Schüler (113)  
 Grundschule (61)  
 Industrie (47)

- ❖ Übersicht zu Studien und verfügbaren Daten der Bildungsforschung
- ❖ Such- und Filtermöglichkeiten
- ❖ Wachsendes Angebot mit mehreren hundert Studien

Image: [www.freephotos.cc](http://www.freephotos.cc) by Skuawk (CC-0).

// Daten finden



## Anmelden

[Benutzername vergessen?](#)

[Passwort vergessen?](#)

Anmelden

[Noch nicht registriert? Hier geht's zur Registrierung.](#)

### Forschende:

- ❖ Einfaches Registrieren und Hochladen der eigenen Forschungsdaten

### VerbundFDB:

- ❖ Erstellen des zentralen Katalogeintrags
- ❖ Weiterleitung zu den zuständigen Datenzentren



informieren



- ❖ Infomaterial
- ❖ individuelle Beratung
- ❖ Workshops
- ❖ Webinare

// Informieren

# Warum Forschungsdatenmanagement?

# Was ist Forschungsdatenmanagement?

---

*Forschungsdatenmanagement ist wie  
die Gesundheitsvorsorge für Ihre  
Forschungsdaten*

Umfasst alle Strategien,  
Prozesse und Maßnahmen

- zum Erhalt der Datenqualität
- zur Interpretierbarkeit von Forschungsergebnissen
- zur Gewährleistung der (Nach-) Nutzbarkeit von Forschungsdaten





# Warum Open Access?

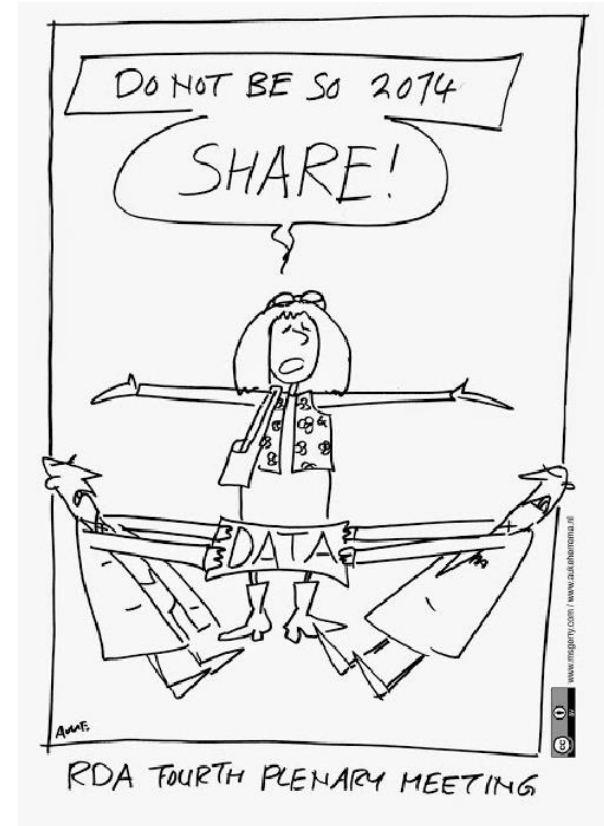
- Förderung der Wissenschaft und Vermeidung von Redundanzen
- Bestmögliche Nutzung öffentlicher Gelder
- Sicherung guter wissenschaftlicher Praxis
- Steigerung der Reputation



# Förderauflagen zu Open Access

*„sollten **Forschungsdaten** (...) verfügbar gemacht werden (...), die eine sinnvolle Nach- und **Weiternutzung durch Dritte ermöglicht.**“ (DFG 2015)*

*„Die Antragsteller verpflichten sich, (...) gewonnenen **Daten** nach Abschluss des Projektes (...) **zur Verfügung zu stellen** (...) **ein eigenes Forschungsdatenmanagement betreiben.**“ (BMBF-Förderung BiSS, 12.01.2015)*



# Ziele des FDM

---

## 1. Qualitätssicherung

lesbare, authentische  
Daten; Vermeidung von  
Datenverlusten

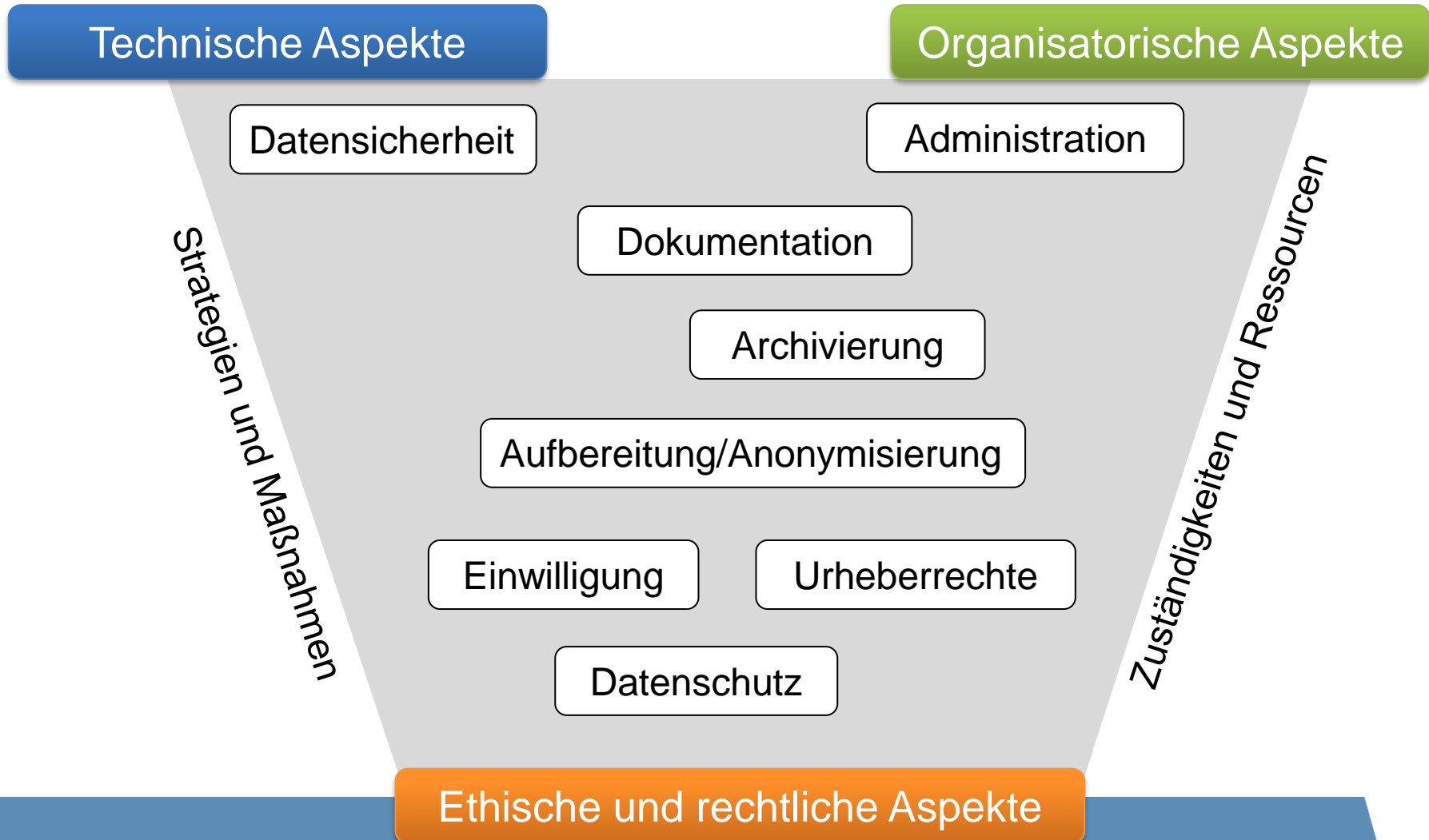
## 2. Replizierbarkeit

Verständlichkeit;  
Reproduzierbarkeit der  
Forschungsergebnisse  
ermöglichen; Nachweis  
der Validität

## 3. Nachnutzbarkeit

Nutzung der Daten in  
neuen Kontexten;  
Beantwortung neuer  
Forschungsfragen

# Bereiche des Forschungsdatenmanagements



# Dokumentation



# Beispiel 1: Mehrdeutige Labels

	Name	Typ	Spalten...	De...	Beschreibung	Werte	Fehlend	Spalten	Ausrichtung	Maß	Rolle
81	f3nc09	Numerisch	1	0	AIDS	{0, -}...	Keine	8	Rechts	Unbekannt	Eingabe
82	f3nc10	Numerisch	1	0	Hepatitis					Unbekannt	Eingabe
83	f3nc11	Numerisch	1	0	BSE					Unbekannt	Eingabe
84	f3nc12	Numerisch	1	0	Diabetes, Zucker					Unbekannt	Eingabe
85	f3nc13	Numerisch	1	0	Milzbrand					Unbekannt	Eingabe
86	f3nc14	Numerisch	1	0	TBC, Tuberkulose					Unbekannt	Eingabe
87	f3nc15	Numerisch	1	0	Hirnschlag, Schlaganfall					Unbekannt	Eingabe
88	f3nc16	Numerisch	1	0	Multiple Sklerose					Unbekannt	Eingabe
89	f3nc17	Numerisch	1	0	Alzheimer					Unbekannt	Eingabe
90	f3nc18	Numerisch	1	0	Fettleibigkeit, Übergewicht					Unbekannt	Eingabe
91	f3nc19	Numerisch	1	0	typ. Kinderkrankheiten					Unbekannt	Eingabe
92	f3nc20	Numerisch	1	0	Gehirnhautentzündung					Unbekannt	Eingabe
93	f3nc21	Numerisch	1	0	Hautkrankheiten					Unbekannt	Eingabe
94	f3nc22	Numerisch	1	0	Gefäßkrankheiten					Unbekannt	Eingabe
95	f3nc23	Numerisch	1	0	Lungen-, Atemwegserkrankungen					Unbekannt	Eingabe
96	f3nc24	Numerisch	1	0	Erkrankungen der inneren Organe	{0, -}...	Keine	8	Rechts	Unbekannt	Eingabe
97	f3nc25	Numerisch	1	0	Knochen-, Gelenkerkrankungen, Bewegungsapparat	{0, -}...	Keine	8	Rechts	Unbekannt	Eingabe
98	f3nc26	Numerisch	1	0	Stress	{0, -}...	Keine	8	Rechts	Unbekannt	Eingabe
99	f3nc27	Numerisch	1	0	Unfälle jeglicher Art	{0, -}...	Keine	8	Rechts	Unbekannt	Eingabe
100	f3nc28	Numerisch	1	0	allg. Seuchen, Epidemien	{0, -}...	Keine	8	Rechts	Unbekannt	Eingabe
101	f3nc29	Numerisch	1	0	Sonstiges	{0, -}...	Keine	8	Rechts	Unbekannt	Eingabe

Wertbeschriftungen

Wert:

Beschreibung:

0 = "-"

1 = "+"

Hinzufügen

Ändern

Entfernen

Rechtschreibung...

OK Abbrechen Hilfe

Bedeutung der Codes?

Bedeutung der Variablen

# Beispiel 2: Fehlender Codeplan

- Zusätzliche Materialien/Dokumente aus der Erhebungsphase können helfen, oder auch nicht:
- Auszug aus Fragebogen

3. Welches sind Ihrer Meinung nach die größten Probleme, die die Gesundheit der deutschen Bevölkerung bedrohen?  
 NENNUNGEN BITTE IM TEXTFELD NOTIEREN

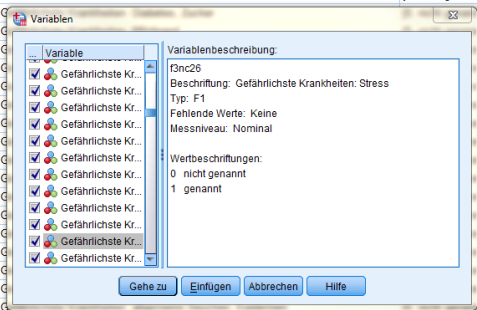
4. Was sind Ihrer Meinung nach in der Bundesrepublik zurzeit die gefährlichsten Krankheiten?  
 NENNUNGEN BITTE IM TEXTFELD NOTIEREN

94	f3nc22	Num...	1	0	Gefäßerkrankungen, Arteriosklerose	{0, -}...
95	f3nc23	Num...	1	0	Lungen-, Atemwegserkrankungen	{0, -}...
96	f3nc24	Num...	1	0	Erkrankungen der inneren Organe	{0, -}...
97	f3nc25	Num...	1	0	Knochen-, Gelenkerkrankungen, Bewegungsapparat	{0, -}...
98	f3nc26	Num...	1	0	Stress	{0, -}...
99	f3nc27	Num...	1	0	Unfälle jeglicher Art	{0, -}...
100	f3nc28	Num...	1	0	allg. Seuchen, Epidemien	{0, -}...
101	f3nc29	Num...	1	0	Sonstiges	{0, -}...
102	f3nc30	Num...	1	0	weiß nicht	{0, -}...
103	f3nc31	Num...	1	0	k.A.	{0, -}...
104	f4c1	Num...	1	0	Bekanntheit von AIDS -Krankheit allgemein	{0, -}...
105	f4c2	Num...	1	0	Bekanntheit von AIDS -Schwächung d. Immunsysteme...	{0, -}...

# Darstellung strukturierter Metadaten (Variablenebene)

## Informationen im Datensatz (SPSS)

Name	Typ	Sp...	De...	Beschriftung	Werte
f3nc09	Numerisch	1	0	Gefährlichste Krankheiten: AIDS	{0, nicht genannt}...
f3nc10	Numerisch	1	0	Gefährlichste Krankheiten: Hepatitis	{0, nicht genannt}...
f3nc11	Numerisch	1	0	Gefährlichste Krankheiten: BSE	{0, nicht genannt}...
f3nc12	Numerisch	1	0		
f3nc13	Numerisch	1	0		
f3nc14	Numerisch	1	0		
f3nc15	Numerisch	1	0		
f3nc16	Numerisch	1	0		
f3nc17	Numerisch	1	0		
f3nc18	Numerisch	1	0		
f3nc19	Numerisch	1	0		
f3nc20	Numerisch	1	0		
f3nc21	Numerisch	1	0		
f3nc22	Numerisch	1	0		
f3nc23	Numerisch	1	0		
f3nc24	Numerisch	1	0		
f3nc25	Numerisch	1	0		
f3nc26	Numerisch	1	0		
f3nc27	Numerisch	1	0		
f3nc28	Numerisch	1	0		
f3nc29	Numerisch	1	0	Gefährlichste Krankheiten: Sonstiges	{0, nicht genannt}...
f3nc30	Numerisch	1	0	Gefährlichste Krankheiten: weiß nicht	{0, nicht genannt}...
f3nc31	Numerisch	1	0	Gefährlichste Krankheiten: keine Angabe	{0, nicht genannt}...
f3zrc1	Numerisch	1	0	Gefährlichste Krankheiten (komprimiert): AIDS	{0, nicht genannt}...



## Informationen im Variablenreport/Codebook

AIDS im öffentlichen Bewusstsein der Bundesrepublik Deutschland 2006  
 GESIS Studiennr. ZA5120, v1.0.0, <http://dx.doi.org/doi:10.4232/1.5120>

f3nc26 - Gefährlichste Krankheiten: Stress

F,4

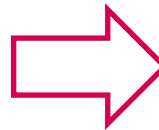
Was sind Ihrer Meinung nach in der Bundesrepublik zurzeit die gefährlichsten Krankheiten?

<NENNUNGEN BITTE IM TEXTFELD NOTIEREN>

F,4\_26 Stress

0 nicht genannt

1 genannt



## Informationen im Fragebogen

- Welches sind Ihrer Meinung nach die größten Probleme, die die Gesundheit der deutschen Bevölkerung bedrohen?

NENNUNGEN BITTE IM TEXTFELD NOTIEREN

- Was sind Ihrer Meinung nach in der Bundesrepublik zurzeit die gefährlichsten Krankheiten?

NENNUNGEN BITTE IM TEXTFELD NOTIEREN

ZA5120, f3nc26: Gefährlichste Krankheiten: Stress (N=3601)

f3nc26

Wert	Ausprägung	Missing	Anzahl	Prozent	Gült.Prozent	Kumul.Prozent
0	nicht genannt		3562	98,9	98,9	98,9
1	genannt		39	1,1	1,1	100,0
	Summe		3601	100,0	100,0	
	Gültige Fälle		3601			

# Darstellung strukturierter Metadaten (Studienebene)

## Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vorschul- und Schulalter (BiKS)

### Laufzeit

01.01.2005 - 31.12.2014

Das Projekt wurde in Form von zwei Längsschnittuntersuchungen durchgeführt. Die Erhebungen im Längsschnitt (BiKS-3-10) fanden sowohl in den Familien als auch in Kindergärten in den Bundesländern Bayern und Hessen statt und begannen im Herbst 2005. Die

### STUDIEN- UND PROJEKTINFORMATIONEN

#### Projektleitung

Artelt, Cordula; Roßbach, Hans-Günther; Weinert, Sabine; Blossfeld, Hans-Joachim

#### Beteiligte Institution(en)

[Fachgruppe Soziologie \(Otto-Friedrich-Universität Bamberg\)](#)  
[Lehrstuhl Psychologie I, Institut für Psychologie \(Otto-Friedrich-Universität Bamberg\)](#)  
[Lehrstuhl für Empirische Bildungsforschung, Institut für Erziehungswissenschaften \(Otto-Friedrich-Universität Bamberg\)](#)

#### Projektwebsite

<http://www.uni-bamberg.de/biks/>

#### Schlagwörter

[Bildungsprozess](#); [Sprachentwicklung](#); [Kognitive Entwicklung](#); [Kompetenzentwicklung](#); [Sekundarstufe I](#); [Schuljahr 01](#); [Schuljahr 02](#); [Schuljahr 03](#); [Schuljahr 04](#); [Übergang Vorschulstufe - Primarstufe](#); [Familie](#); [Schullaufbahnentscheidung](#); [Sekundarstufe I](#); [Lernbedingungen](#); [Lernumgebung](#); [Kompetenzmessung](#)

#### Förderung

Deutsche Forschungsgemeinschaft (DFG); Otto-Friedrich-Universität Bamberg

### TEILSTUDIE: BiKS-3-10 - BILDUNGSPROZESSE, KOMPETENZENTWICKLUNG UND SELEKTIONSENTSCHEIDUNGEN IM VORSCHUL- UND SCHULALTER

#### Laufzeit der Studie

01.01.2005 - 31.12.2014

#### Untersuchungsdesign

Quasi-Experimentelles Design  
 Längsschnitt: Panel

#### Erhebungsmethode(n)

### FORSCHUNGSDATEN UND -INSTRUMENTE

#### BiKS-3-10 - Daten

##### Datenzugang

[DOI: 10.5159/IQB\\_BIKS\\_3\\_10\\_v4](#)

##### Art der Daten

Umfrage- und Aggregatdaten  
 Kompetenz- und Leistungsdaten

##### Archivierende Einrichtung

[Forschungsdatenzentrum am Institut zur Qualitätsentwicklung im Bildungswesen \(FDZ am IQB\)](#)

##### Zugänglichkeit

Beantragung nötig

##### Erhebungszeitraum

01.09.2005 - 30.06.2012

##### Erhebungsverfahren

Interview; Persönliches Interview; Eigenständig auszufüllender Fragebogen; Papierfragebogen; Leistungs- und Kompetenztests; Paper and Pencil Test

# Datenschutz und Informierte Einwilligung



# Rechtmäßigkeit der Verarbeitung

## Artikel 6

### Rechtmäßigkeit der Verarbeitung

- (1) Die Verarbeitung ist nur rechtmäßig, wenn mindestens eine der nachstehenden Bedingungen erfüllt ist:
- a) Die betroffene Person hat ihre Einwilligung zu der Verarbeitung der sie betreffenden personenbezogenen Daten für einen oder mehrere bestimmte Zwecke gegeben;
  - b) die Verarbeitung ist für die Erfüllung eines Vertrags, dessen Vertragspartei die betroffene Person ist, oder zur Durchführung vorvertraglicher Maßnahmen erforderlich, die auf Anfrage der betroffenen Person erfolgen;
  - c) die Verarbeitung ist zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt;
  - d) die Verarbeitung ist erforderlich zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt;
  - e) die Verarbeitung ist für die Erfüllung eines öffentlichen Auftrags erforderlich, der dem Verantwortlichen übertragen wurde;
  - f) die Verarbeitung ist zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person, die den Schutz personenbezogener Daten erfordern, überwiegen, insbesondere dann, wenn es sich bei der betroffenen Person um ein Kind handelt.

**Entscheidend für uns ist die  
freiwillige und informierte Einwilligung.**

(Art. 6 Abs. 1 DSGVO)

# Personenbezogene Daten

---

„... alle Informationen, die sich auf eine **identifizierte oder identifizierbare natürliche Person** (im Folgenden „betroffene Person“) beziehen;

als identifizierbar wird eine natürliche Person angesehen,

- die direkt oder indirekt,
- insbesondere mittels **Zuordnung [zu spezieller Angabe, z.B. Namen] oder**
- **zu einem oder mehreren besonderen Merkmalen**, die Ausdruck der physischen, physiologischen, etc. Identität dieser natürlichen Person sind, identifiziert werden kann;“

(Art. 4 Nr. 1 DSGVO)

# Besondere Kategorien personenbezogener Daten

---

- Rassistische und ethnische Herkunft
- Politische Meinungen
- Religiöse oder weltanschauliche Überzeugungen
- Gewerkschaftszugehörigkeit
- Genetische Daten
- Biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person
- Gesundheitsdaten
- Daten zum Sexualleben oder der sexuellen Orientierung

(Art. 9 DSGVO)

# Die informierte Einwilligung

---

„jede **freiwillig** für den **bestimmten Fall**, in **informierter Weise** und unmissverständlich abgegebene **Willensbekundung** in Form einer **Erklärung**“  
(Art. 4.11 DSGVO)

- Muss freiwillig abgegeben werden
- Ist an (einen) bestimmte(n) Zweck(e) gebunden
- Muss in informierter Art und Weise abgegeben werden
- Ist eine ausdrückliche Willenserklärung



# Fallstricke in der Informierten Einwilligung

---

- Informationen zum Projekt fehlen.
- “Ihre Daten verbleiben im Projekt und werden nicht weitergegeben.” oder “Die Daten werden nur innerhalb des Forschungsprojekts genutzt.”
- “Durch Ihre Teilnahme stimmen Sie zu.”
- Stabile Kontaktadresse angeben!



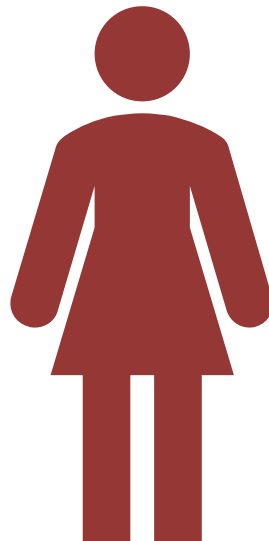
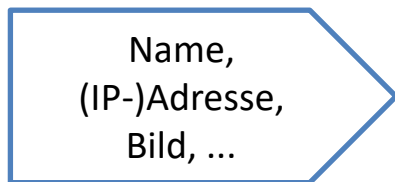
# Weitere Fallstricke

---

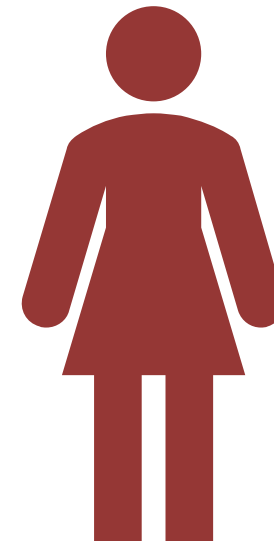
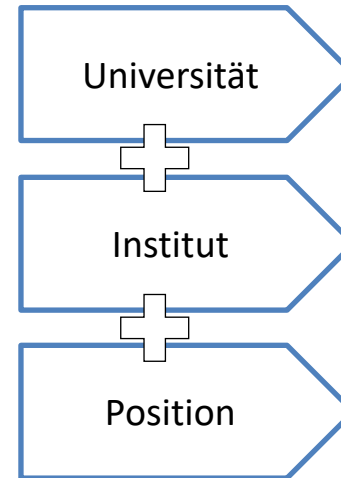
- Keine Trennung zwischen Adress- und Befragungsdaten: “Ihre Daten werden anonymisiert.”
- Gegensätzliche Angaben: “Ihre Adressdaten werden gelöscht.”, aber “Wir werden Sie nochmals kontaktieren.”
- Auf angemessene Sprache und Umfang achten.
- Archivierung und Nachnutzung der Daten sollten schon in der Einwilligung erwähnt werden
- **Beispiele für Formulierungen:** <https://www.forschungsdaten-bildung.de/einwilligung>

# Direkte und indirekte Identifikatoren

## Direkt



## Indirekt



# Anonymisierung der erhobenen Daten

Kritische Variable	Risikierte Folge	Maßnahme
<b>Geburtsdatum</b>	Kann Identifizierung vereinfachen	Aggregation zu Jahr oder Monat und Jahr
<b>Kinder im Haushalt</b>	Hohe Kinderzahl kann Alleinstellungsmerkmal sein	Abschneidegrenze, z.B. "6 Kinder oder mehr"
<b>Herkunftsland</b>	Seltene Einwanderungsland kann Alleinstellungsmerkmal sein	Seltene Länder gruppieren in Regionen
<b>Familienstand</b>	"eingetragene Lebenspartnerschaft" identifiziert Homosexualität	Variablenausprägung mit "verheiratet" zusammenlegen
...		

# Die Archivierung von Daten

# Dateneingangsprüfung im Verbund

- **Rechtliche Anforderungen ✓**
  - Keine urheberrechtlichen Verletzungen und geklärte Nutzungsrechte
  - Ausreichende Anonymisierung personenbezogener Daten
  - informierte Einwilligung zur Verarbeitung und Weitergabe der Daten liegt vor
  
- **Technische Anforderungen ✓**
  - Daten liegen in gängigen Formaten vor bzw. Software ist vorhanden
  - Prüfung auf Schadcodes/Viren
  
- **Dokumentarische Anforderungen ✓**
  - Plausibilität und Konsistenz der Daten
  - Vollständigkeit der Dokumentation
  - Dokumentation und Daten kongruent

# Suche nach geeignetem Repository

<https://www.re3data.org/>



## Mapping the global repository landscape

International collaboration has become very common in the production of scientific knowledge (Coccia and Wang, 2016; Monastersky & Van Noorden, 2019)

## How open are repositories in re3data?

Openness may sound self-evident, but in fact it can mean different things, even within the Open Science community (Pasquetto, Sands & Borgman,

## Reviewing the subject classification in re3data

When looking for a suitable repository, the discipline(s) a repository caters to is one of the most important pieces of information. The re3data Metadata

# Danke für die Aufmerksamkeit!

## Noch Fragen?

Per Mail:

[forschungsdaten-bildung@gesis.org](mailto:forschungsdaten-bildung@gesis.org) oder [marcus.eisentraut@gesis.org](mailto:marcus.eisentraut@gesis.org)

Per Telefon: **069 / 24708-300 (DIPF)** oder **0221 / 47694-473 (M.Eisentraut)**



# Quellenhinweis

- » Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](#).
- » Folien 19-21, 23-26 und 29 sind in abgewandelter Form übernommen von:
- » Perry, Anja, Mozygemba, Kati, Weber, Judith, Linares, Jimena, Eberhardt, Jan, Mosbrugger, Volker, Taffner, Julian, & Tschink, Daniel. (2022, May 25). Workshop zu Forschungsdatenmanagement für die BMBF-Forschungsinitiative zum Erhalt der Artenvielfalt (FEEdA). Zenodo. <https://doi.org/10.5281/zenodo.6579410> (CC-BY)