Institut zur Qualitätsentwicklung im Bildungswesen

Forschungsdatenzentrum

DATEN TEILEN & OPEN SCIENCE IN DER EMPIRISCHEN BILDUNGSFORSCHUNG

CC BY

Dr. Aleksander Kocaj &
Christin Rüdiger M.A.

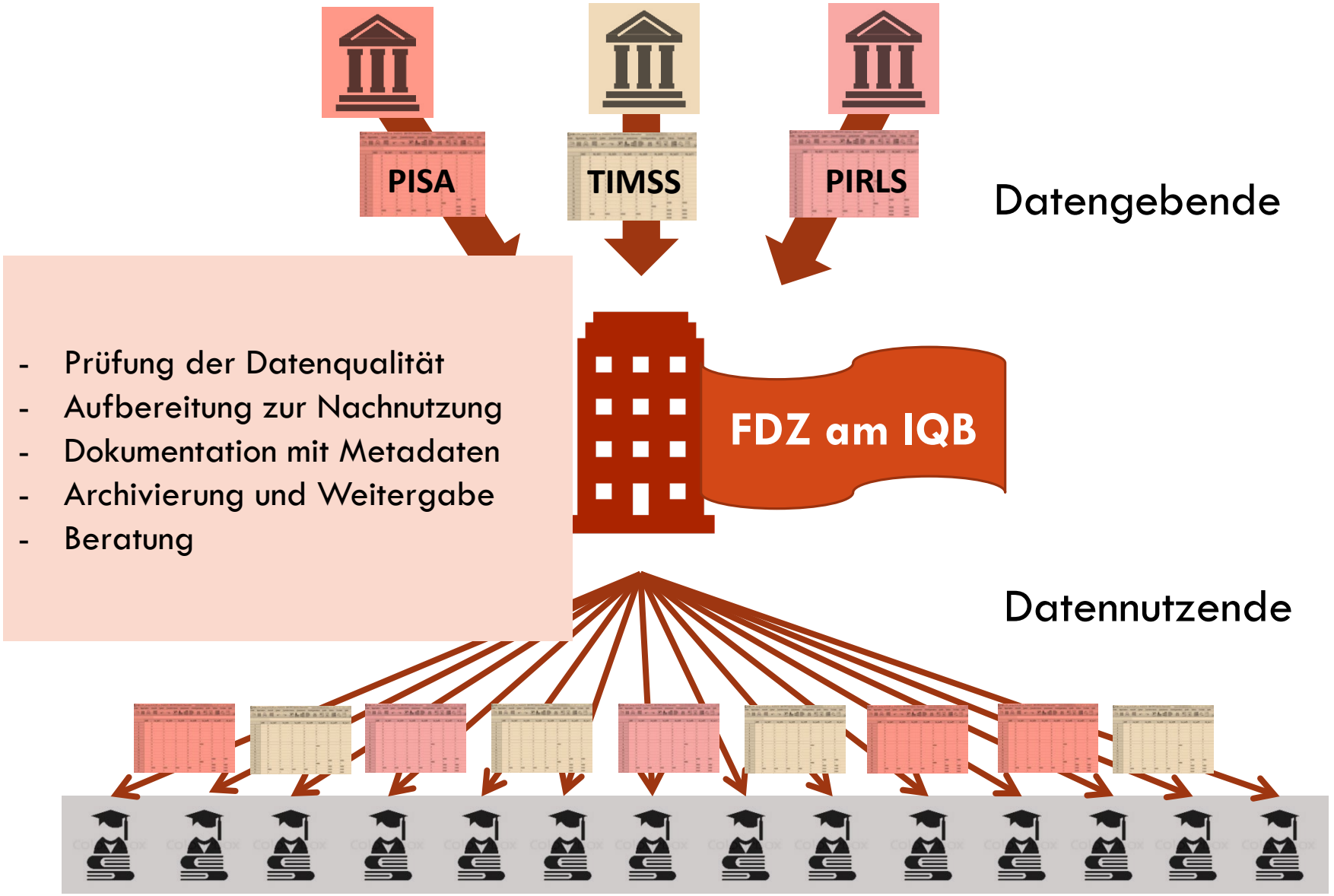
Themen des Workshops

- Replikationskrise
- Überblick über Open Science Praktiken und Akteure
- Datenteilen – Wieso, wie, wo?
- Datendokumentation
- Datenschutz und sensible Daten
- Automatisierte Skalenhandbucherstellung
- Synthetische Daten erzeugen
- Reproduzierbarer Code

FORSCHUNGSDATENZENTRUM AM IQB

- Teil des Zentrums für internationale Bildungsvergleichsstudien (ZIB)
- Partner im Verbundprojekt Forschungsdaten Bildung
- akkreditiert vom Rat für Sozial- und Wirtschaftsdaten
- Core Trust Seal - Zertifizierung

WAS MACHT DAS FDZ AM IQB?



DATENSÄTZE AM FDZ AM IQB

- insgesamt derzeit 60 Studien verfügbar
- Fokus auf quantitative Daten mit Kompetenzmessungen im Bildungsbereich

LISA 6

Lernergebnisse an beruflichen und allgemeinbildenden Gymnasien in Schleswig-Holstein

KAT-HS

Adaptive University Exams

BILWISS

ICILS 2013 IEA

TIMSS 2011

BiSpra Bildungssprachliche Kompetenzen

PISA 2015

OECD PISA Deutschland

StEG Studie zur Entwicklung von Ganztagsschulen

DESI

MARKUS

BiKS

BiLieF

IQB-Bildungstrend 2015 informieren Sie sich

KuL

IGLU 2016

Ascot Technologie-orientierte Kompetenzmessung in der Berufsbildung Technology-based Assessment of Skills and Competencies in VET

Ländervergleich 2011 informieren Sie sich

Professionswissen in den Naturwissenschaften

prowin

ELEMENT

WAS MACHT DER VerbundFDB?

VerbundFDB
forschungsdaten-bildung.de

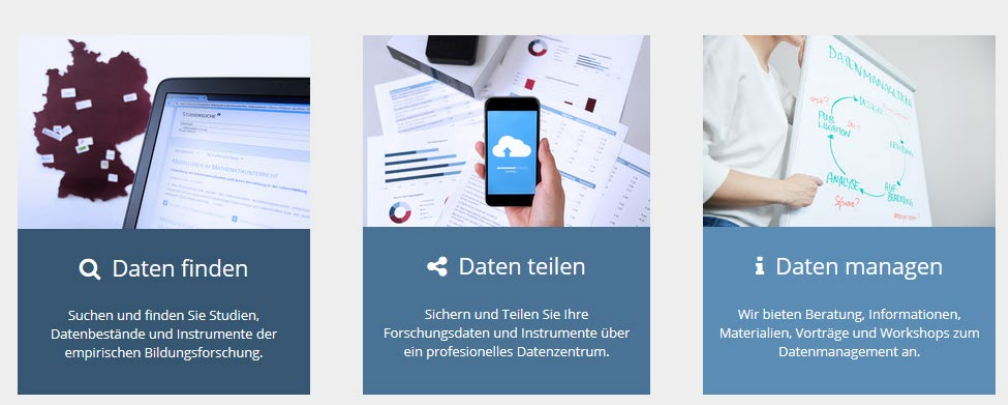
Meta-Services

Qualiservice data sharing *
bibb Bundesinstitut für Berufsbildung Forschungsdatenzentrum – FDZ *
ApaeK
DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation *
DIE
MERCATOR INSTITUT für Sprachförderung und Deutsch als Zweitsprache
gesis Leibniz-Institut für Sozialwissenschaften *
fdz.DZHW Forschungsdatenzentrum *
IQB Institut zur Qualitätsentwicklung im Bildungswesen *
LIfBi LEIBNIZ-INSTITUT FÜR BILDUNGSVERLÄUFE e.V. *
SOEP *
DJI Deutsches Jugendinstitut *
leibniz-psychology.org *

* akkreditiert beim RatSWD

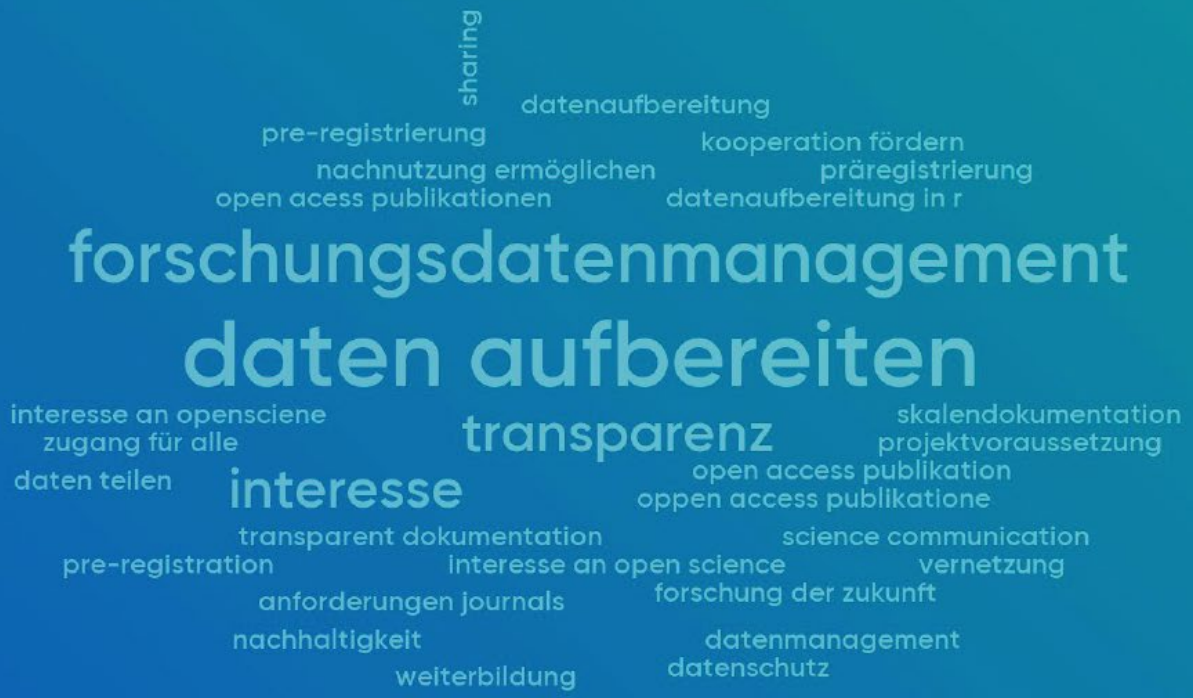
Zentrale Services

(1) Archivierung von Daten
(2) Zentraler Zugang zu Forschungsdaten
(3) Angebote zum Forschungsdatenmanagement



Warum nehmen Sie am Workshop teil?

Mentimeter



27

DIE REPLIKATIONS-KRISE ALS STARTPUNKT FÜR VERÄNDERUNGEN IN DER FORSCHUNGS-PRAXIS

General Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Psychological Science
22(11) 1359–1366
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797611417632
http://pss.sagepub.com
SAGE

Joseph P. Simmons[1], Leif D. Nelson[2], and
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of

Journal of Personality and Social Psychology
2012, Vol. 103, No. 6, 933–948
© 2012 American Psychological Association
0022-3514/12/$12.00 DOI: 10.1037/a0029709

Correcting the Past: Failures to Replicate Psi

Robyn A. LeBoeuf
University of Florida

Joseph P. Simmons
University of Pennsylvania

3,289), we replicate the procedure of Experiments 8 and 9 from Bem (2011),
cilitation of recall. We failed to replicate that finding.
ation attempts of these experiments and find that the
n 0. We discuss some reasons for differences between
Bem (2011).

grees of freedom, meta-analysis

Journal of Personality and Social Psychology
2011, Vol. 100, No. 3, 426–432
© 2011 American Psychological Association
0022-3514/11/$12.00 DOI: 10.1037/a0022790

Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)

Eric–Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, and Han L. J. van der Maas
University of Amsterdam

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

REPLIKATIONSKRISE?

BEISPIEL FÜR MISSLUNGENE REPLIKATION: FACIAL FEEDBACK HYPOTHESIS

Originalstudie:

Pen-in-mouth paradigm (Strack, Martin & Stepper, 1988)

Koordinierte Replikation:

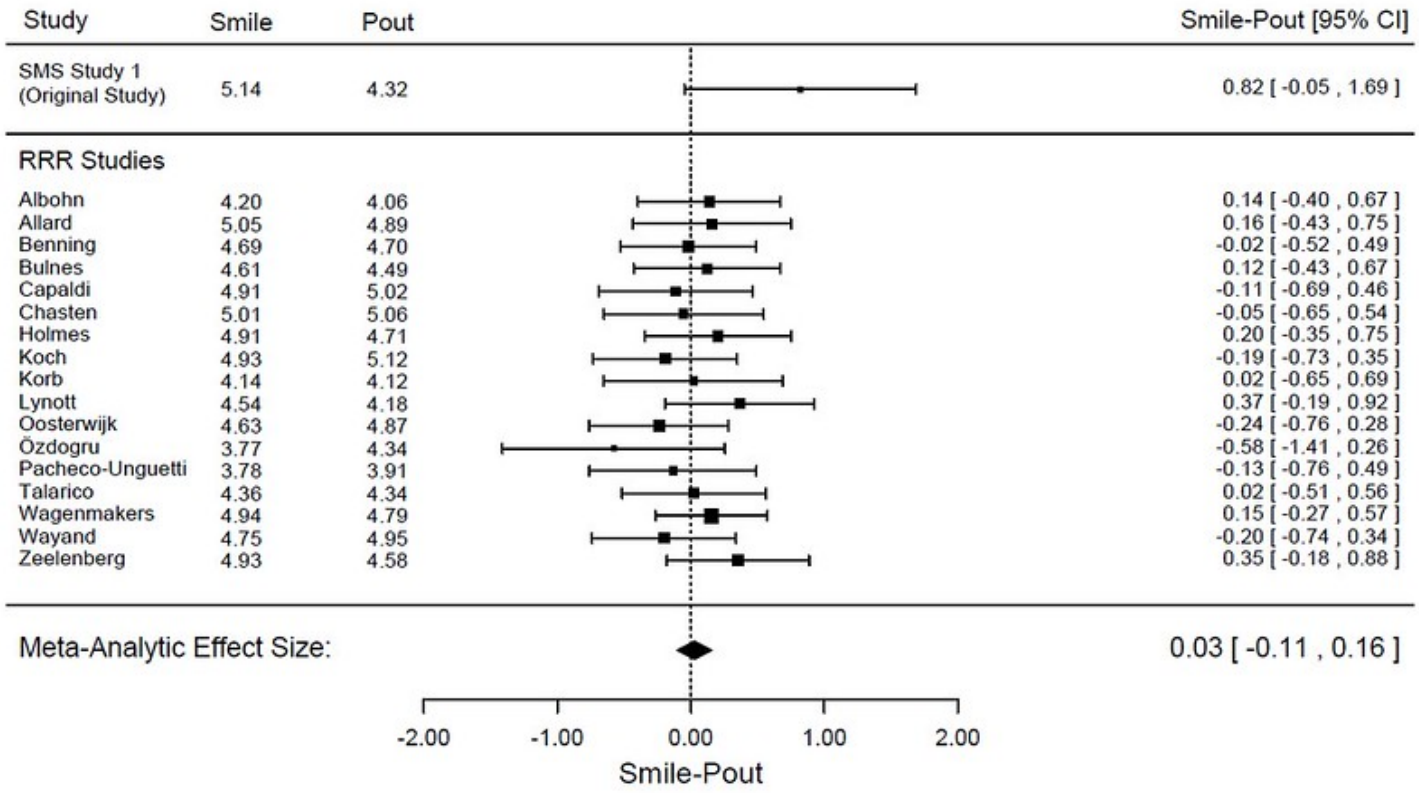
Registered Replication Report (Wagenmakers et al., 2016)

Gronau, CC-BY, accessible under http://tinyurl.com/zm7p9l7

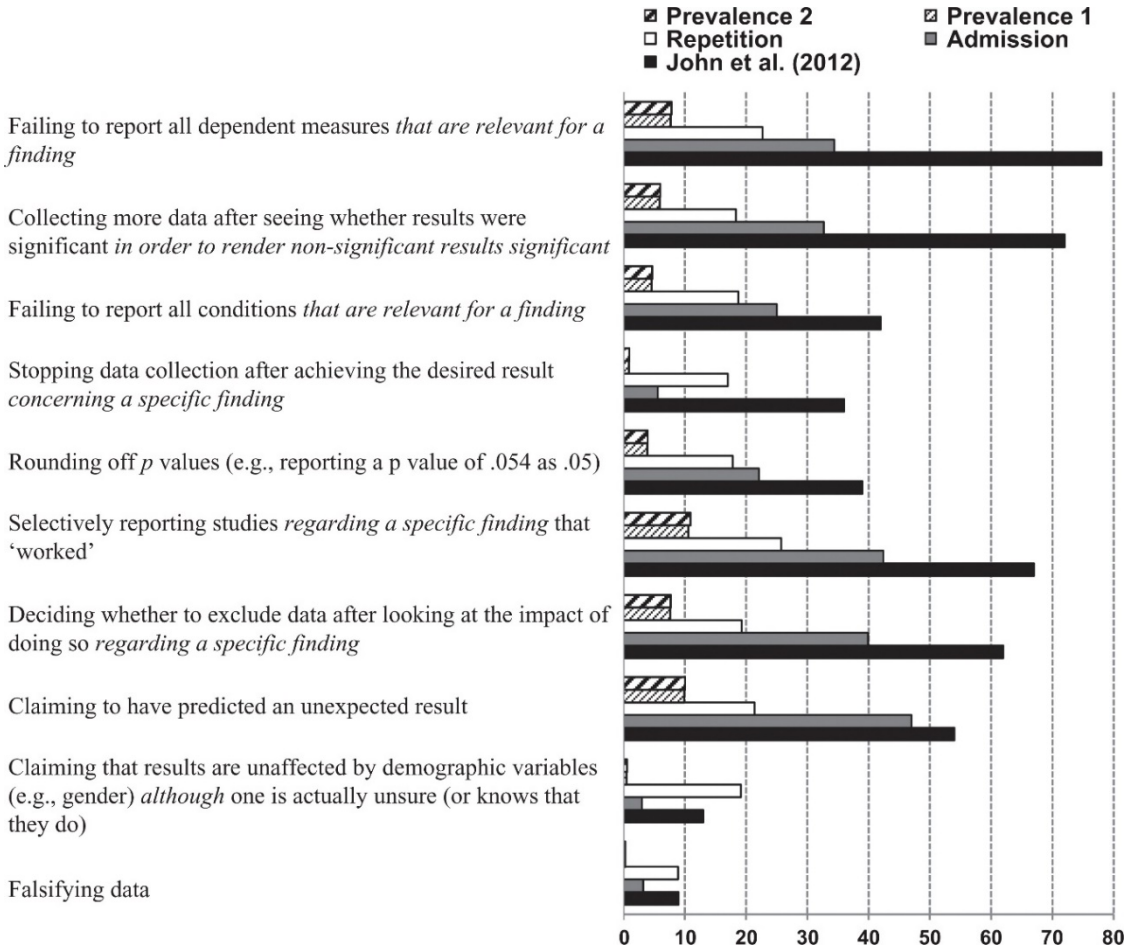
https://doi.org/10.1177/1745691616674458

BEISPIEL FÜR MISSLUNGENE REPLIKATION

Registered Replication Report (Wagenmakers et al., 2016, S.923)



https://doi.org/10.1177/1745691616674458

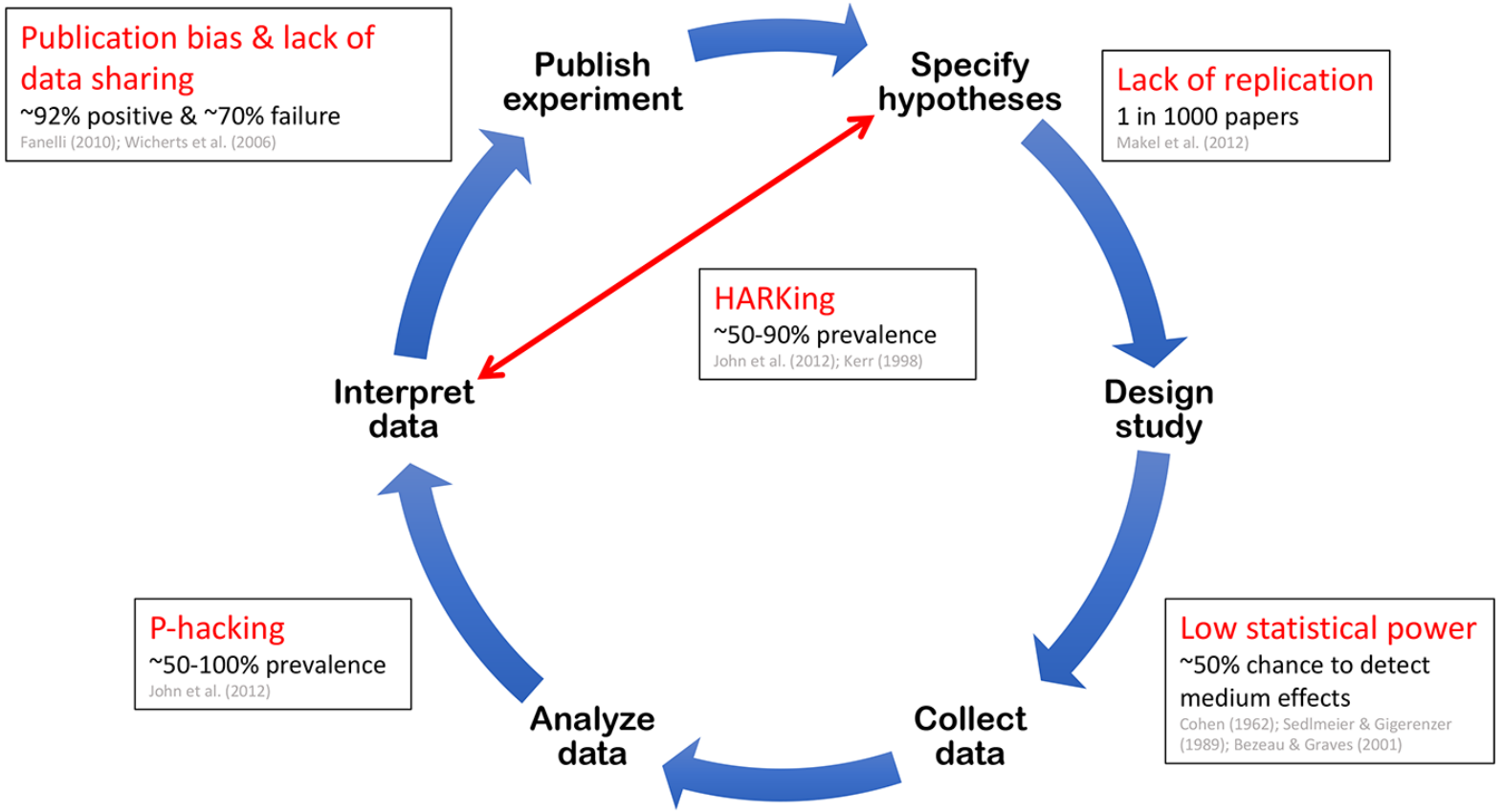


Prävalenz fragwürdiger Forschungspraktiken

Figure 2. Prevalence indices (shaded bars) derived from admission rates of respondents committing questionable research practices at least once (gray bars) and repetition frequency (white bars), compared to the original John et al. (2012) data (black bars). Modified item wordings appear in italics.

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. Social *Psychological and Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

WARUM IST DIE REPLIKATIONSRATE SO GERING?



Munafò et al. (2017, p. 2): http://dx.doi.org/10.1038/s41562-016-0021

WARUM IST DIE REPLIKATIONSRATE SO GERING?

REPLIKATION IN DER BILDUNGSFORSCHUNG

Rosenthal & Jacobson (1964):https://doi.org/10.1007/BF02322211

Duncan et al. (2014): http://dx.doi.org/10.1037/a0037996

Raudenbush (1984): https://doi.org/10.1037/0022-0663.76.1.85

Journal of Educational Psychology
1984, Vol. 76, No. 1, 85–97

Copyright 1984 by the
American Psychological Association, Inc.

Magnitude of Teacher Expectancy Effects on Pupil IQ as a Function of the Credibility of Expectancy Induction: A Synthesis of Findings From 18 Experiments

Stephen W. Raudenbush
Graduate School of Education, Harvard University

This inquiry employs meta-analysis (Glass, 1977) to account for variability in the outcomes of experiments testing the effects of teacher expectancy on pupil IQ. The tenuous process of expectancy induction, wherein researchers supply teachers with information designed to elevate their expectancies for children actually selected at random, is viewed as the Achilles' heel of Pygmalion (Rosenthal & Jacobson, 1968) experiments. It was hypothesized that the better teachers know their pupils at the time of expectancy induction, the smaller the treatment effect would be. The data strongly supported this hypothesis. Hypotheses that the type of IQ test (group vs. individual) and type of test administrator (aware vs. blind to expectancy-inducing information) influence experimental results were not supported. The hypothesis that expectancy effects are larger for children in Grades 1 and 2 than for children in Grades 3–6 was supported. However, surprisingly, significant effects reappeared at Grade 7. Theoretical implications and questions for future meta-analytic research are discussed.

REPLIKATION IN DER BILDUNGSFORSCHUNG

Makel & Plucker (2014, S.311):
https://doi.org/10.3102/0013189X14545513

- 0.13% aller veröffentlichten Studien sind Replikationen (221 von 164589 Studien)
- diese Replikationsrate ist 8mal geringer als in der Psychologie
- in den letzten 20 Jahren zeigt sich ein Anstieg der Replikationsrate

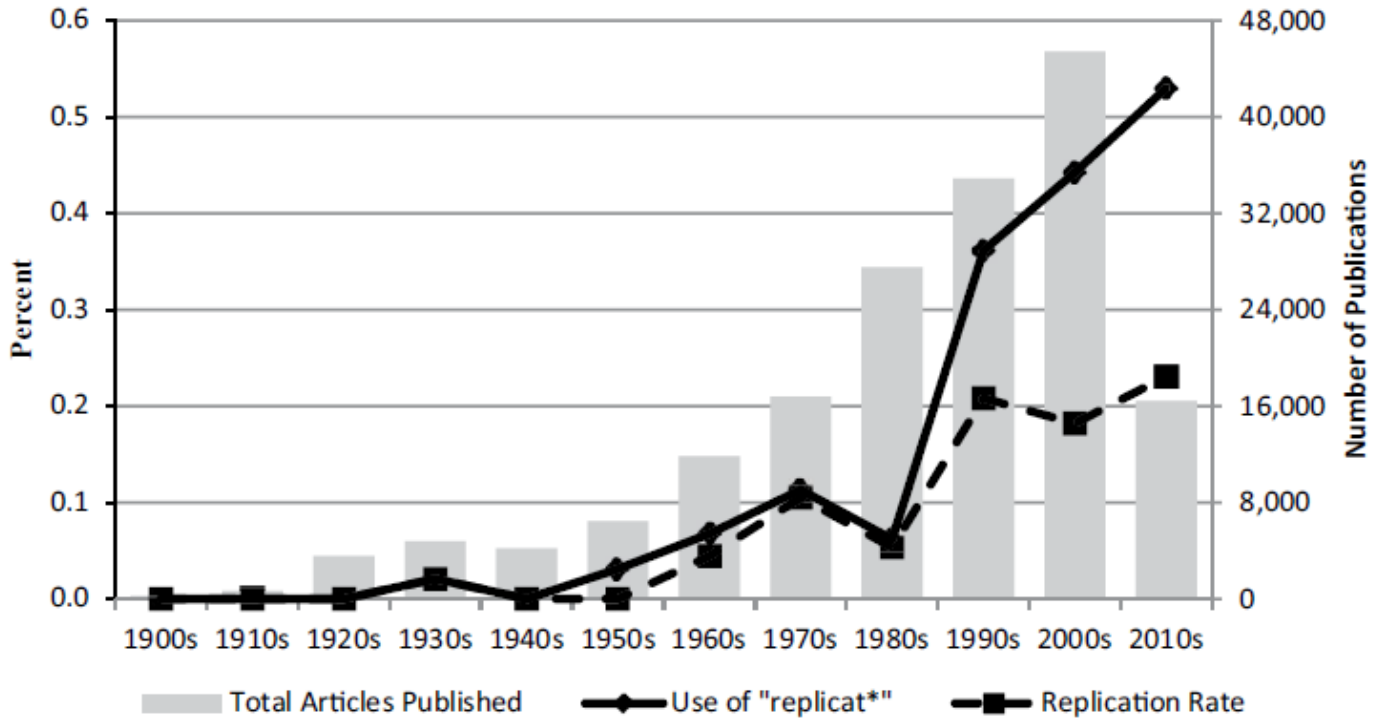


FIGURE 1. *Replication rate in education journals*

REPLIKATION IN DER BILDUNGSFORSCHUNG

Makel & Plucker (2014, S. 311):
https://doi.org/10.3102/0013189X14545513

- insgesamt berichteten 67.4% aller Replikationsstudien eine erfolgreiche Replikation
- exakte Replikationen sind häufiger erfolgreich als konzeptuelle Replikationen
- Erfolgsrate ist geringer, wenn Replikation von unterschied. Forscher*innengruppe durchgeführt wird

Table 2
Replication Results in Education Journals

	Replication Outcome			
Replication Type	**Success**	**Failure**	**Mixed**	**Total**
Direct	45 (71.4%)	11 (17.5%)	7 (11.1%)	63 (28.5%)
Conceptual	101 (66%)	18 (11.8%)	34 (22.2%)	153 (69.2%)
Mixed	3 (60%)	0 (0%)	2 (40%)	5 (2.3%)
Total	149 (67.4%)	29 (13.1%)	43 (19.5%)	221 (100%)
Replication Authorship	**Success**	**Failure**	**Mixed**	**Total**
Same Publication as Original	47 (88.7%)	2 (3.8%)	4 (7.5%9)	53 (24%)
Same Authors New Publication	48 (70.6%)	6 (8.8%)	14 (20.6%)	68 (30.8%)
All Unique Authors	54 (54%)	21 (21%)	25 (25%)	100 (45.2%)

AK

REPLIKATION IN DER BILDUNGSFORSCHUNG

Rost & Bienefeld (2019):
https://doi.org/10.1024/1010-0652/a000253

- Zeitschrift für Pädagogische Psychologie ermutigt zur Einreichung von Replikationsstudien – bei Erscheinen des Beitrags (24.10.2019) wurde noch keine einzige eingereicht!
- Replikationsstudien sollen nun bevorzugt publiziert werden
- "If a study is worth doing at all, it's worth doing twice" (Lindsay & Ehrenberg, 1993, S. 226)

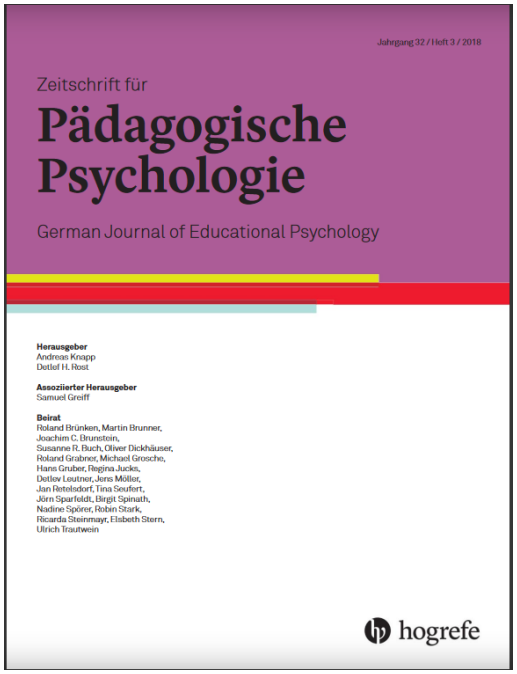
Editorial

Nicht replizieren: publizieren!?

Detlef H. Rost[1,2] und Marc Bienefeld[3]

[1]Southwest University Chongqing, Faculty of Psychology, Chongqing, P.R. China
[2]Philipps-Universität Marburg, Fachbereich Psychologie, Marburg, Deutschland
[3]Universität Bielefeld, Fakultät für Erziehungswissenschaft, Bielefeld, Deutschland

Zusammenfassung: In der (Pädagogischen) Psychologie sind Replikationsstudien bislang extrem seltene Ausnahmen. Dieser Artikel legt dar, dass und warum Wiederholungsstudien unentbehrlich sind. Weiterhin wird der Frage nachgegangen, warum – trotz des enormen Mehrwerts – nahezu keine Replikationen publiziert werden und warum viele „Ergebnisse" der psychologischen Forschung nicht replizierbar sind. Dass es sich bei diesen Sachverhalten nicht um Vermutungen handelt, wird durch vorliegende Untersuchungen belegt. Die Ursachen dafür liegen in verschiedenen – teilweise voneinander abhängigen – Ebenen des Wissenschaftssystems: die verbreitete – aber abwegige – Ansicht, „statistische Signifikanz" indiziere auch die Wahrscheinlichkeit, einen Befund replizieren zu können; die Verwechslung von „statistisch signifikant" mit relevant; die Unsitte, getestete Untersuchungshypothesen erst im Nachhinein (*ex post*), also in Kenntnis der Resultate einer Studie, aufgestellt zu haben, aber in der Publikation als theoretisch abgeleiteten Ausgangspunkt (d. h. *a priori* formuliert) auszugeben; die α-Fehler-Inflationierung durch multiple statistische Signifikanztestungen; das exklusive Berichten von Ergebnissen, welche die Forschungshypothesen stützen, verbunden mit dem Unterschlagen abweichender Befunde; mangelnde Konstruktvalidität der verwendeten Messinstrumente; Lug und Betrug in der Wissenschaft; die Geringschätzung von Replikationen durch Zeitschriftenherausgeber, Gutachter und Drittmittelgeber. All das führt dazu, dass fast ausschließlich „statistisch signifikante" und „neue" Ergebnisse veröffentlicht werden und falsche Theorien persistieren. Als Gegenmaßnahmen werden beispielhaft genannt: eine großzügige finanzielle Förderung von Replikationsprojekten und ihrer Publikation; die nachdrückliche gutachterliche Befürwortung der Veröffentlichung methodisch adäquater Wiederholungsstudien; die Bereitschaft von Fachzeitschriften, dafür genug Platz bereitzustellen; die Anerkennung des großen wissenschaftlichen Werts von Wiederholungsstudien, auch in Berufungsverfahren. Daraus ergibt sich, dass mit den aufgezeigten Möglichkeiten und Forderungen zur Etablierung und Förderung von Replikationsstudien unterschiedliche Adressaten parallel angesprochen werden müssen. Nachhaltige Veränderungen sind allerdings nur erreichbar, wenn die einzelnen Akteure (Forscher; Gutachter; Zeitschriftenherausgeber; Berufungskommissionen; Drittmittelgeber) ihre individuelle Verantwortung anerkennen und entsprechende Taten folgen lassen.



8 MODULAR STANDARDS

CITATION STANDARDS Cite shared data to incentivize their publication	**DATA TRANSPARENCY** Disclose, require, or verify shared data
ANALYTICAL METHODS TRANSPARENCY Disclose, require, or verify shared code	**RESEARCH MATERIALS TRANSPARENCY** Disclose, require, or verify shared materials
DESIGN AND ANALYSIS TRANSPARENCY Sets standards for research design disclosures	**PREREGISTRATION OF STUDIES** Specification of study details before data collection
PREREGISTRATION OF ANALYSIS PLANS Specification of analytical details before data collection	**REPLICATION** Encourages publication of replication studies

ACROSS 3 TIERS

DISCLOSURE:
The article must disclose whether or not materials are available.

REQUIREMENT:
The article must share materials when possible.

VERIFICATION:
Third party must verify that the standard is being met.

OPEN SCIENCE ALS LÖSUNGSANSATZ

ÜBERBLICK ÜBER OPEN SCIENCE PRAKTIKEN

"For education researchers, summarizing and reporting how we conduct our investigations is among our most important professional responsibilities."

van der Zee & Reich (2018)

DOI: 10.1177/2332858418787466

Was sind Ihre bisherigen Erfahrung mit Open Science Praktiken?

Mentimeter

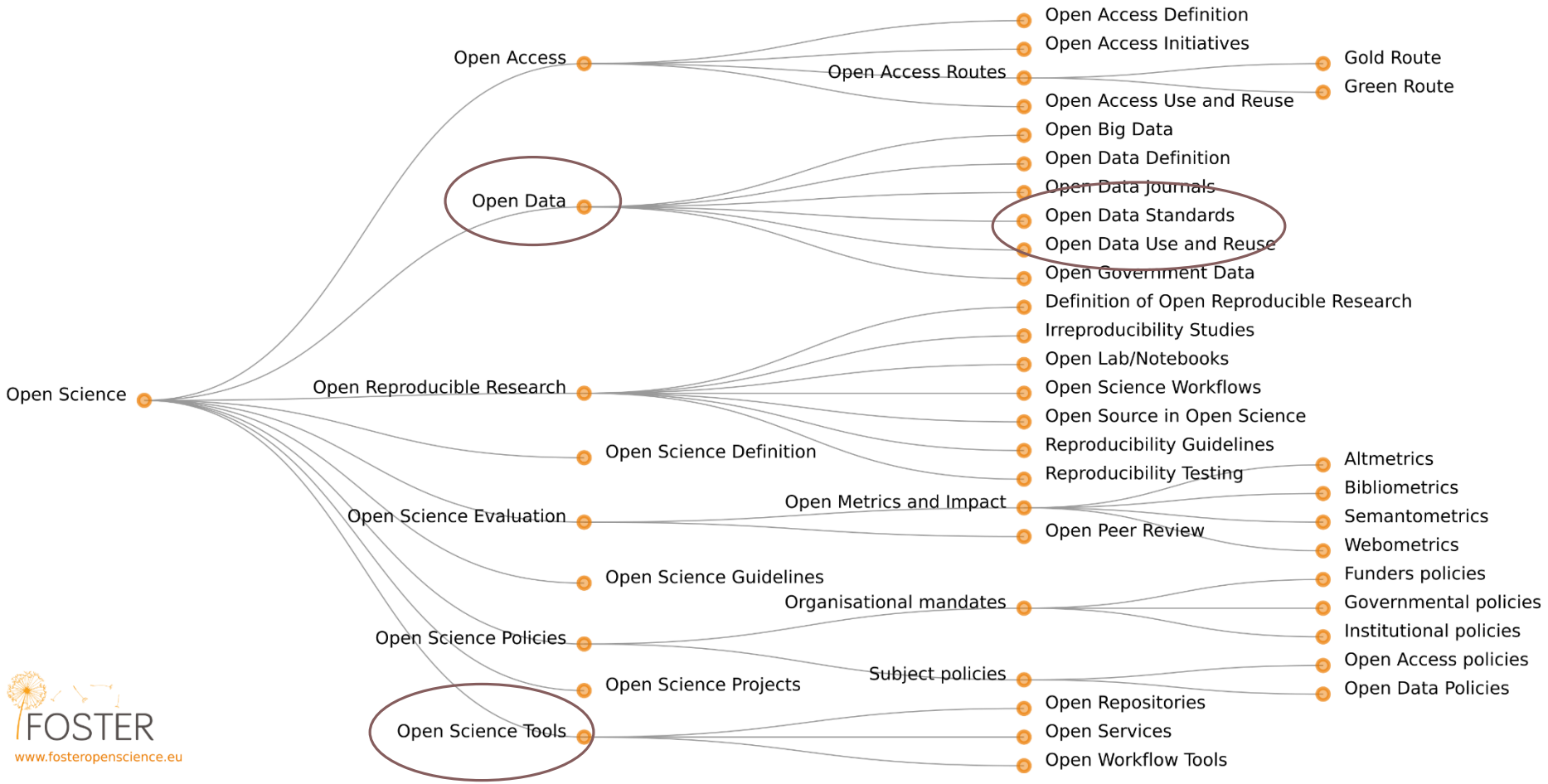
27

OPEN SCIENCE?

"Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods. In a nutshell, Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks (Vicente-Sáez & Martínez-Fuentes 2018)."

https://open-science-training-handbook.gitbook.io/book/open-science-basics/open-concepts-and-principles

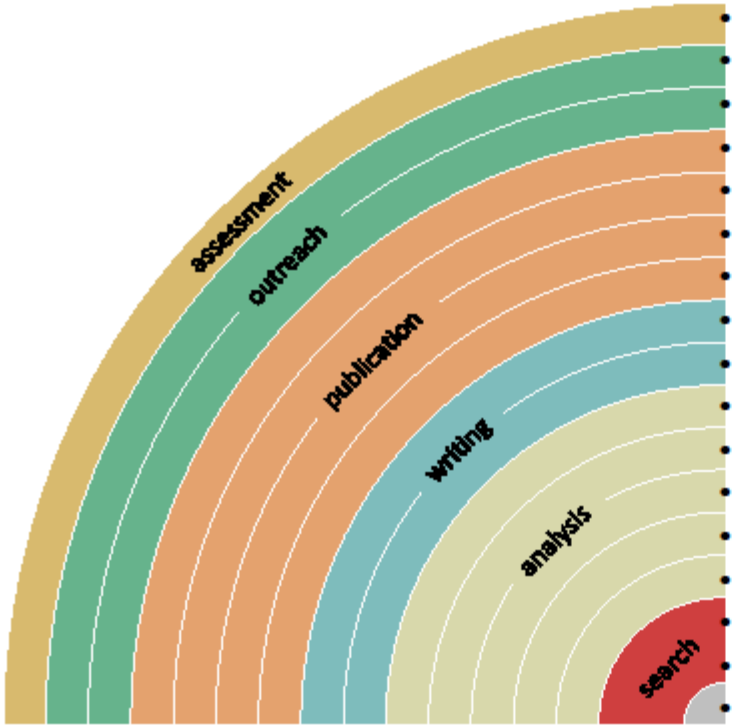
Open Science Taxonomy



FOSTER
www.fosteropenscience.eu

https://www.fosteropenscience.eu/themes/fosterstrap/images/taxonomies/os_taxonomy.png

You can make your workflow more open by …



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO

Bianca Kramer & Jeroen Bosman https://101innovations.wordpress.com

DOI: 10.5281/zenodo.1147025

https://zenodo.org/record/1147025

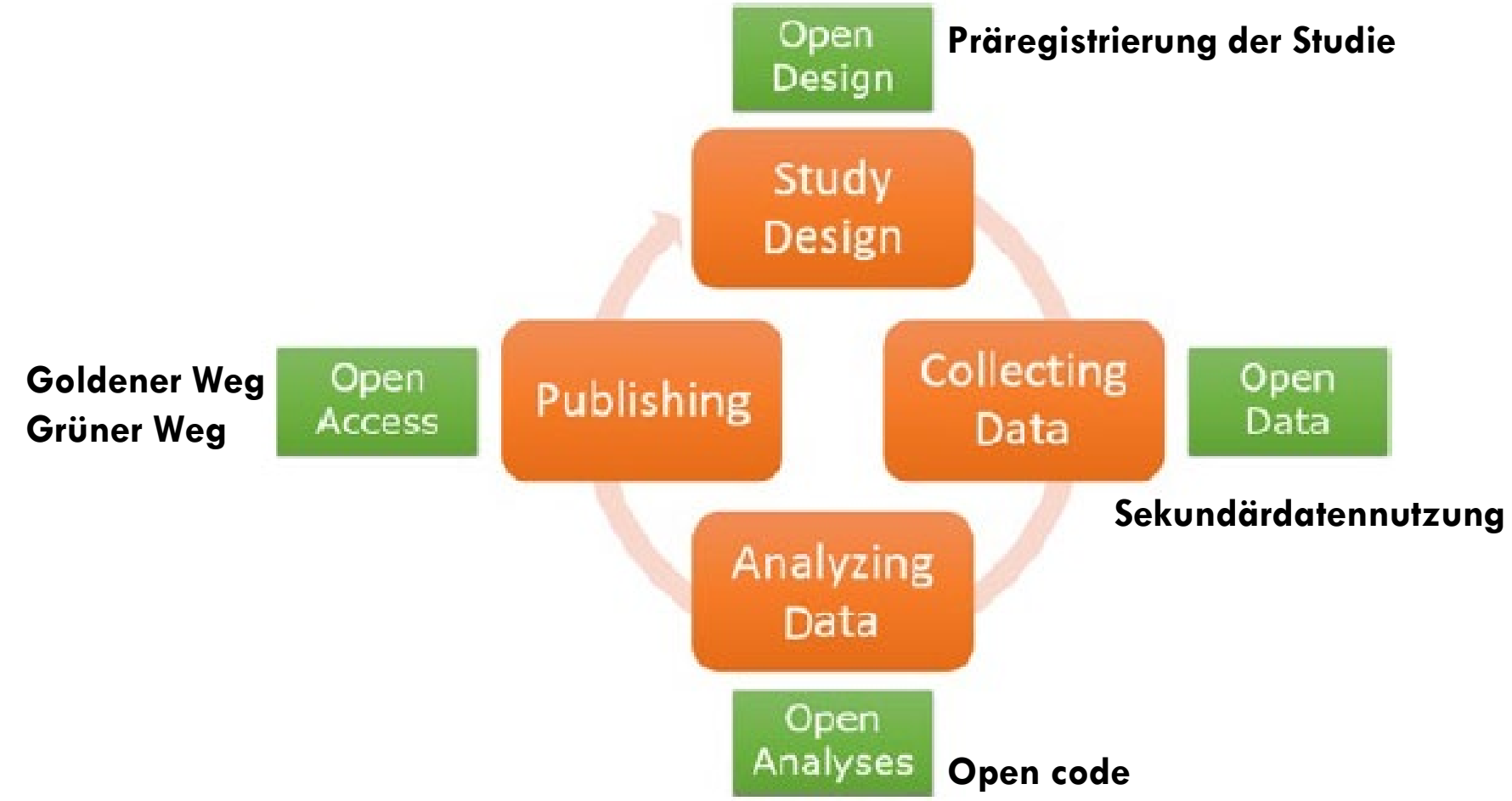


FIGURE 1. *The open research cycle.*

van der Zee & Reich (2018) DOI: 10.1177/2332858418787466

For internal communication, we are using a Google group. If you're interested, feel free to join! https://groups.google.com/d/forum/nosi-de

Existing Open Science Initiatives

In chronological order according to founding date, including the contact person:

- Ludwig-Maximilians-Universität München (Felix Schönbrodt felix@nicebread.de)
- Max-Planck-Institut für Kognitions- und Neurowissenschaften (Lieneke Janssen janssen@cbs.mpg.de)
- Universität Koblenz-Landau (Tanja Könen koenen@uni-landau.de)
- FernUniversität Hagen (Andreas Glöckner andreas.gloeckner@fernuni-hagen.de)
- Universität Münster (Mitja Back mitja.back@wwu.de)
- Universität Göttingen, GOSSIP (Tanja Gerlach gerlach@uni-goettingen.de)
- Universität Göttingen, Open Science MeetUp (Birgit Schmidt bschmidt@sub.uni-goettingen.de, Julika Mimkes mimkes@sub.uni-goettingen.de)
- Universität Wien (Martin Voracek martin.voracek@univie.ac.at)
- Humboldt-Universität Berlin (Kai Horstmann horstmak@cms.hu-berlin.de)
- Friedrich-Alexander-Universität Erlangen-Nürnberg (Oliver Schultheiss oliver.schultheiss@fau.de)
- Universität Leipzig (Stefan Schmukle schmukle@uni-leipzig.de und Julia Rohrer julia.rohrer@posteo.de)
- Universität Marburg: OSIUM (Open Science Initiative University Marburg) (José Alanis und Peer Herholz)
- Universität Marburg: Fachbereich Psychologie (Erik Mueller erik.mueller@staff.uni-marburg.de)
- Universität Köln: Department Psychologie (Christoph Stahl christoph.stahl@uni-koeln.de)
- Universität Zürich: Psychologisches Institut (Johannes Ullrich j.ullrich@psychologie.uzh.ch)
- Goethe-Universität Frankfurt: Open Science Initiative (Julia Beitner beitner@psych.uni-frankfurt.de, Christian Fiebach fiebach@psych.uni-frankfurt.de)
- Technische Universität Dresden: Open Science Initiative der Fakultät Psychologie (Stefan Scherbaum stefan.scherbaum@tu-dresden.de)
- Universität Erfurt: "Erfurter Open Science Initiative" (EFOSI) (Thomas Lauer thomas.lauer@uni-erfurt.de & Stefanie Lindow stefanie.lindow@uni-erfurt.de)
- Mannheim: Mannheim Open Science Meetup (Jana Berkessel jana.berkessel@uni-mannheim.de>)
- Universität zu Lübeck (Finn Lübber f.luebber@uni-luebeck.de und Lena Rademacher lena.rademacher@uni-luebeck.de)

WER IST DABEI?

OSF

https://osf.io/tbkzh/wiki/home/

WER IST DABEI?

Übersicht des Publikationsoutputs der Datennutzenden des FDZ am IQB

Hier finden Sie eine Übersicht der Publikationen, die mit Daten, die am FDZ am IQB vorliegen, erstellt wurden. Diese Bibliographie umfasst Forschungsartikel sowie Bücher.

Wenn Sie Daten des FDZ am IQB in Publikationen verwendet haben, helfen Sie uns bitte, diese Bibliographie zu vervollständigen, indem Sie uns Zitation/en sowie eine Kopie der Artikel schicken (entweder in Papierform oder elektronisch).

Bibliographie

Anzahl: **166** (Stand: 12.02.2021)

2020

Bach, M. & Fischer, M. (2020). Understanding the response to high- stakes incentives in primary education. *IZA Discussion Paper Series*, (13845), 1-88.

Helm, F., Arens, A. K. & Möller, J. (2020). Perceived teacher unfairness and student motivation in math and German: An application of the generalized internal/external frame of reference model. *Learning and Individual Differences, 81,* 101891. https://doi.org/10.1016/j.lindif.2020.101891

Hülshoff, A. (2020). *Fachspezifisches Interesse und wahrgenommene Unterrichtsqualität in heterogenen Lerngruppen. Theoretische Überlegungen und empirische Analysen am Beispiel des Englischunterrichts in der Jahrgangsstufe 9* - Dissertation. Universität Koblenz-Landau, Koblenz-Landau.

Kocaj, A., Jansen, M., Kuhl, P. & Stanat, P. (2020). Zusammenhänge der Klassenkomposition an Förderschulen und allgemeinen Schulen mit schulischen Kompetenzen, akademischem Selbstkonzept und Interesse [Relations between classroom composition at special education schools and regular schools with individual school achievement, academic self-concept, and interest]. In C. Gresch, P. Kuhl, M. Grosche, C. Sälzer & P. Stanat (Hrsg.), *Schüler*innen mit sonderpädagogischem Förderbedarf in Schulleistungserhebungen* (Bd. 39, S. 213-262). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27608-9_8

Kuhl, P., Kocaj, A. & Stanat, P. (2020). Zusammenhänge zwischen einem gemeinsamen Unterricht und kognitiven und non-kognitiven Outcomes von Kindern ohne sonderpädagogischen Förderbedarf. *Zeitschrift für Pädagogische Psychologie*, 1-26. https://doi.org/10.1024/1010-0652/a000283

https://www.iqb.hu-berlin.de/fdz/Publikationsoutput/

WER IST DABEI?

The German Reproducibility Network (GRN) is a **cross-disciplinary consortium that aims to increase trustworthiness and transparency of scientific research** by investigating and encouraging the factors that contribute to robust research. We promote training activities and disseminate best practices, conduct and support meta-scientific research, and work with stakeholders to ensure coordination of efforts.

Please click on any logo to view the respective organization's representatives in GRN.

DGPs Deutsche Gesellschaft für Psychologie

HELMHOLTZ Open Science

HELMHOLTZAI ARTIFICIAL INTELLIGENCE COOPERATION UNIT

ZBW Leibniz Informationszentrum Wirtschaft / Leibniz Information Centre for Economics

OSC LMU Open Science Center

NOSI Netzwerk der Open Science Initiativen

BIH QUEST Transforming Biomedical Research

Berlin University Alliance
Freie Universität Berlin
Humboldt-Universität zu Berlin
Technische Universität Berlin
Charité – Universitätsmedizin Berlin

RN DE German Reproducibility Network

https://reproducibilitynetwork.de/

WARUM SOLLTE ICH OPEN SCIENCE PRAKTIKEN KENNEN UND ANWENDEN?

Die eigene Arbeit wird besser auffindbar, nachvollziehbarer und vertrauenswürdiger. Das kann zu mehr Zitationen und Renommee führen.

Fragwürdige Forschungspraktiken werden reduziert. Das kann das Vertrauen in wissenschaftliche Erkenntnisse insgesamt stärken.

Forschung wird aus öffentlichen Mitteln finanziert. Forschende sollten der Öffentlichkeit daher möglichst viele Informationen zugänglich machen.

WARUM SOLLTE ICH OPEN SCIENCE PRAKTIKEN KENNEN UND ANWENDEN?

Open Science bedeutet, einen Wandel im wissenschaftlichen Arbeiten aktiv umzusetzen. Das ist nicht immer einfach oder sofort möglich.

Daher sollten Sie sich zunächst einen Aspekt davon aussuchen, der Ihnen vielversprechend erscheint.

UMSETZBARKEIT?

Leicht

- Open Access publizieren (wenn Mittel vorhanden)
- Präsentationsfolien lizensieren und teilen
- Daten nachnutzen

Mittel

- Daten und Materialien an ein Repositum übergeben
- Open Science Policy für die eigene Arbeitsgruppe
- Datenmanagmentplan erstellen

Fortgeschritten

- Reproduzierbarkeits- und Robustheitstests
- ManyLabs-Projekte
- Open Code, Open Software und Open Material

BREAKOUT SESSIONS

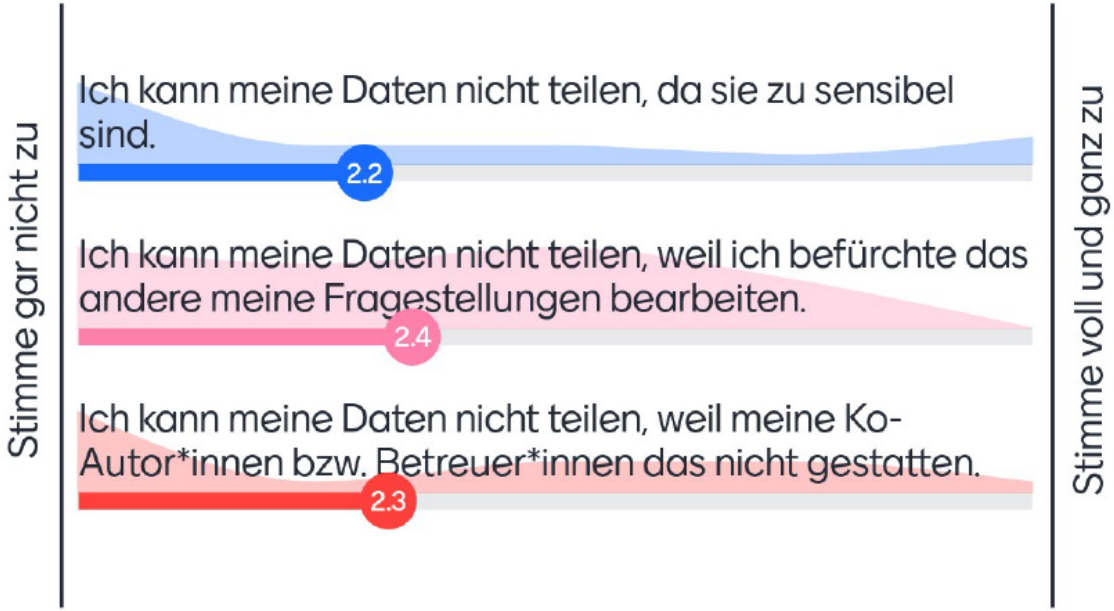
Denke Sie an Ihr aktuelles Forschungsprojekt:

Welchen Open Science Praktiken ließen sich leicht, welche schwieriger umsetzen?

DATEN TEILEN – WIESO?

Welche Vorbehalte haben/hätten Sie beim Teilen von Daten?

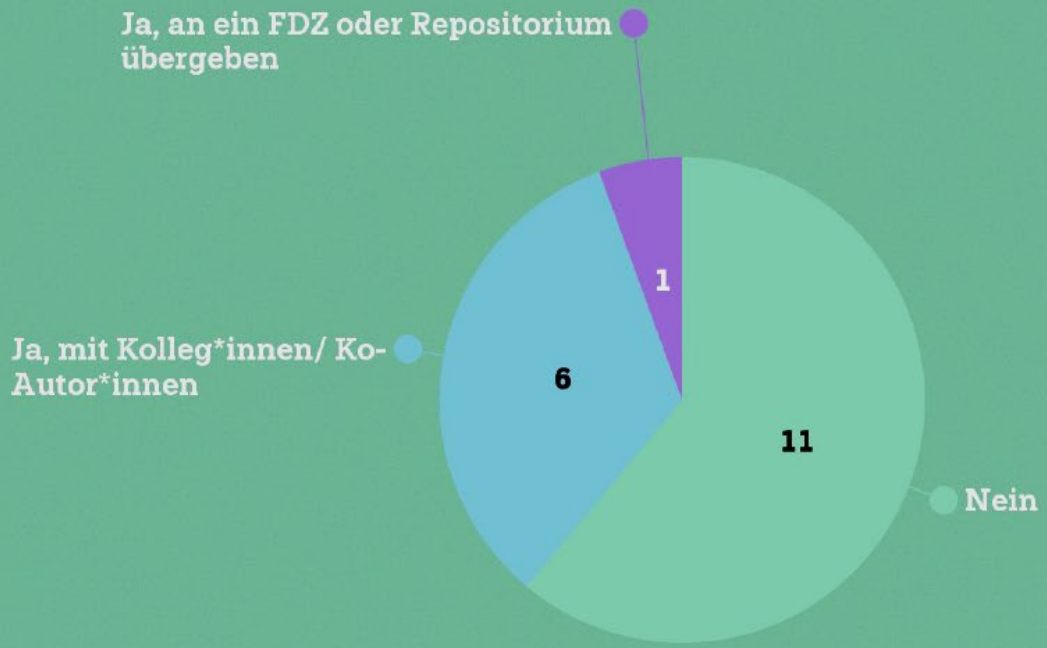
Mentimeter



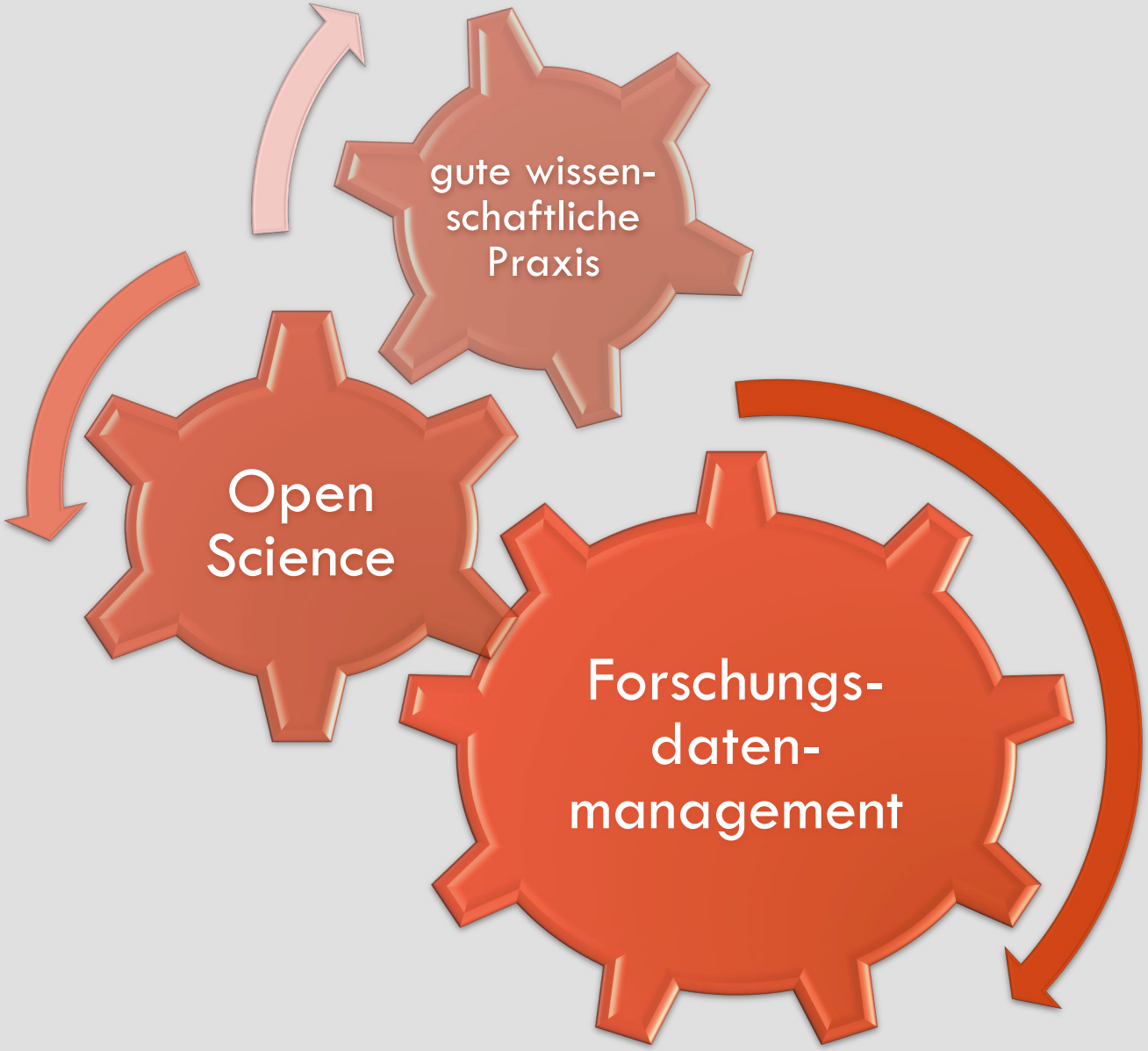
17

Haben Sie selbst bereits Daten geteilt?

Mentimeter



18


gute wissen-
schaftliche
Praxis
Open
Science
Forschungs-
daten-
management

OPEN DATA?

"Open research data is data that can be freely accessed, reused, remixed and redistributed, for academic research and teaching purposes and beyond. Ideally, open data have no restrictions on reuse or redistribution, and are appropriately licensed as such. In exceptional cases, e.g. to protect the identity of human subjects, special or limited restrictions of access are set. Openly sharing data exposes it to inspection, forming the basis for research verification and reproducibility, and opens up a pathway to wider collaboration. At most, open data may be subject to the requirement to attribute and sharealike (see the Open Data Handbook)."

https://open-science-training-handbook.gitbook.io/book/open-science-basics/open-research-data-and-materials

OPEN DATA ≠ DATA SHARING

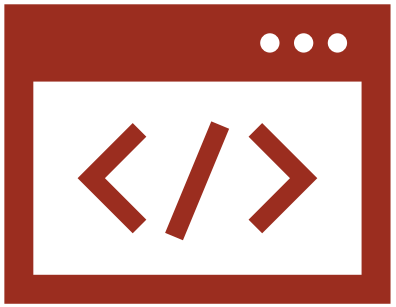
DATENARTEN IN DER EMPIRISCHEN BILDUNGSFORSCHUNG

- Kompetenzdaten
- Befragungsdaten
- Beobachtungsdaten
- Interviewtranskripte
- Video- und Audiodaten
- Prozess-Daten z.B. Reaktionszeit, Eye-Tracking-Daten, Klickverhalten
- sonstige

DATEN TEILEN – WIESO?

Weil man man eine Anfrage bekommt von …

- Ko-Autor*innen
- Journals & Reviewer*innen
- Forschenden, die Ergebnisse replizieren möchten



DATEN TEILEN – WIESO?

Weil die Mittelgeber eines Projektes (z.B. DFG und BMBF) fordern, dass Daten zur Nachnutzung an ein Repositorium übergeben werden müssen

Bundesministerium für Bildung und Forschung

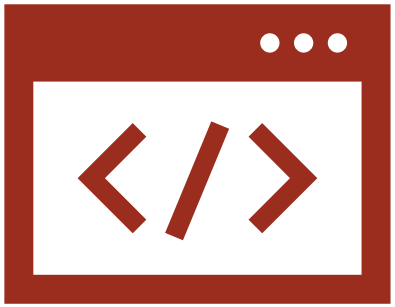
This Photo by Unknown Author is licensed under CC BY-SA

DFG Deutsche Forschungsgemeinschaft

Siehe auch

Übersicht des VerbundFDB
https://www.forschungsdaten-bildung.de/daten-teilen-motive?la=de

Sowie Handreichung vom FDM Thüringen
https://zenodo.org/record/3739190



DATEN TEILEN – WIESO?

Weil der Datensatz, den man mühevoll erarbeitet hat

langzeitarchiviert & nachgenutzt & zitiert

werden soll.

Data Management Horror Stories

DATEN TEILEN – WIESO NICHT?

https://youtu.be/N2zK3sAtr-4

DATEN TEILEN – WIESO NICHT?

Nachdem er seine Dateien geteilt hatte, war sein Team nicht mehr gut auf ihn zu sprechen.

Auflösung

Ein Forscher arbeitete in einem 6-köpfigen Team, um eine Enzyklopädie im Umfang von 800.000 Wörtern zu schreiben. Das Team hatte nach längerer Zeit entschieden, dass alle Arbeiten auf einem gemeinsamen Arbeitsbereich geteilt werden sollen. Hierfür lag schon ein Server vor, der von einigen Mitgliedern kooperativ verwendet wurde für die Projektdaten. Also verschob er das Resultat seiner Arbeit auf den gemeinsamen Speicher und plötzlich waren die anderen sauer.

Das Problem bestand darin, dass sowohl die eigenen Dateien des Forschers, als auch die bereits vorhandenen Dateien der anderen Teammitglieder generisch benannt wurden. Der Verzicht auf eine explizite, zweckgebundene Namenskonvention führte dazu, dass beim Kopieren der lokalen Dateien auf den gemeinsamen Speicher viele Dateien überschrieben wurden. Der Forscher dachte sich nicht viel dabei, da man die alten Dateien sicher durch ein Backup wiederherstellen könnte – nur leider lag dies mehr als 1 Monat zurück und das aktuelle Backup hat durch schlechtes Timing nur die neuen Dateien gesichert.

Die Geschichte verdeutlicht zwei wesentliche Punkte des FDM: Zum einen sollten Organisationsstrukturen und Namenskonventionen festgelegt werden. So können z.B. Datum, Thema und Autor/-in Teil des Dateinamens sein. Zum anderen sollte Wert auf ein gründliches Versionierungs- und Backup-System gelegt werden.

Quelle:

» persönliche Kommunikation

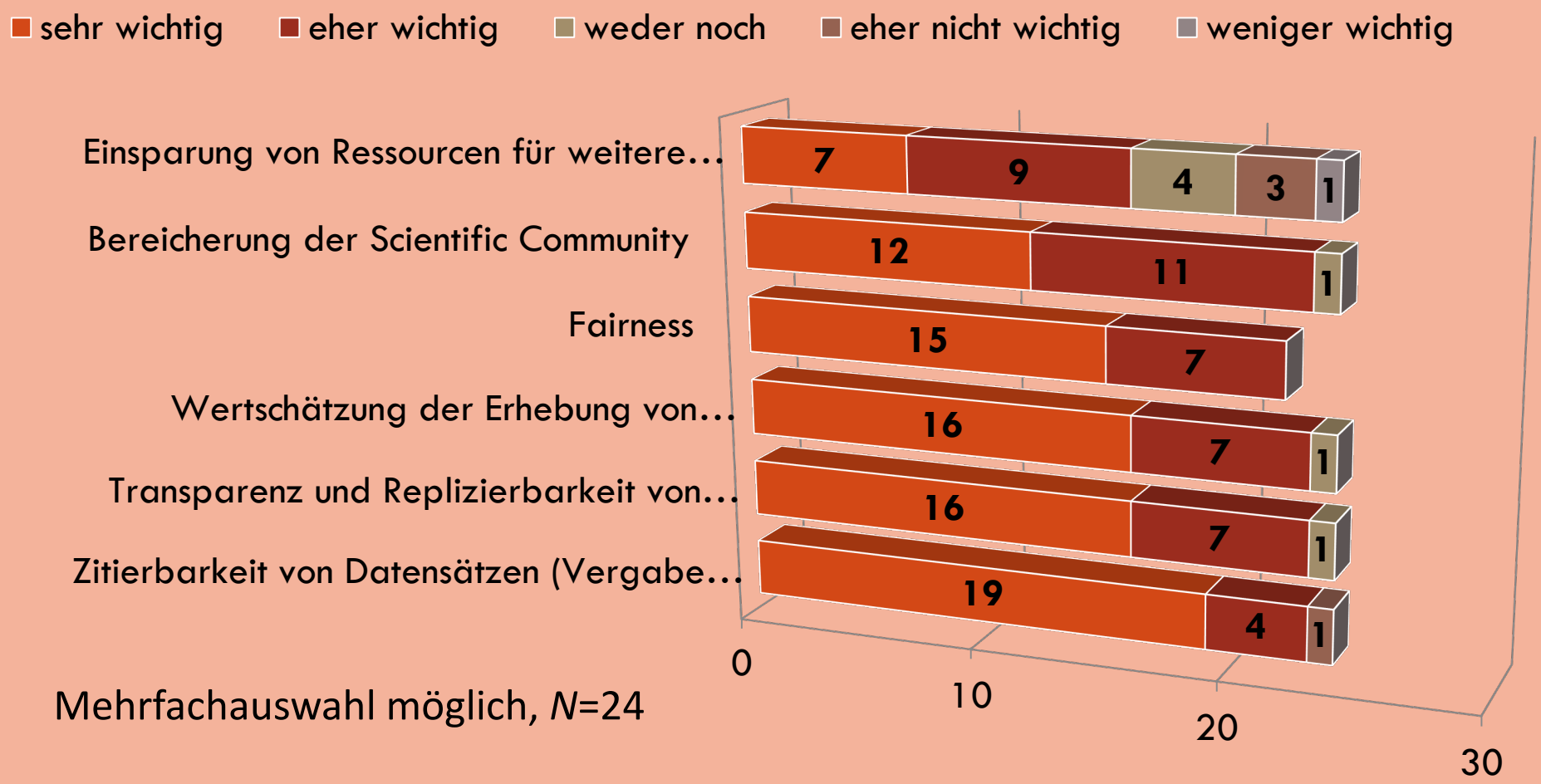
https://forschungsdaten-thueringen.de/geschichten/articles/sharing-is-caring-ger.html

VORTEILE BEIM TEILEN VON DATEN

- **Entlastung der Zielpopulationen** = Erhöhung der Teilnahmebereitschaft (in der empirischen Bildungsforschung sehr wichtig, da z. B. Schulen sehr belastet sind mit Erhebungen)
- **zusätzliche Mittel für Datenaufbereitung und Dokumentation** = in Projektanträgen werden zunehmend Datenmanagementpläne gefordert
- **interdisziplinäre Forschung** = Perspektive und Methodik unterschiedlicher Disziplinen auf den Datensatz

DATENGEBENDENBEFRAGUNG DES FDZ AM IQB (2018)

„Wie wichtig sind folgende Aspekte beim Data Sharing für Sie persönlich?“



Mehrfachauswahl möglich, *N*=24

VORTEILE BEIM NACHNUTZEN VON DATEN

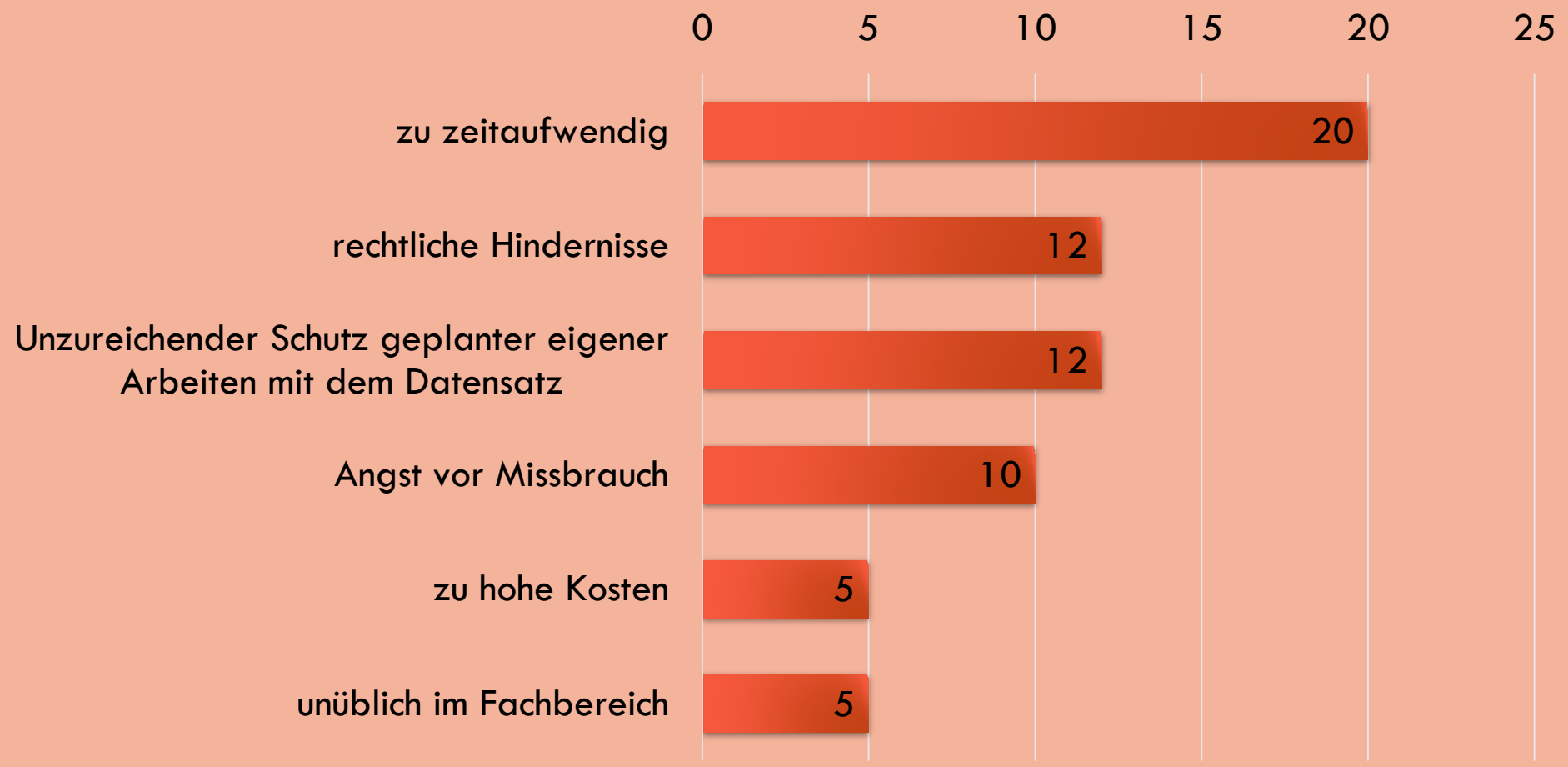
- **Vermeidung unnötiger Erhebungen** = Kostenreduktion & Zeitersparnis
- **umfassende Nutzung von hochwertigen Daten** = in Projektanträgen teilweise Nachweis gefordert, dass man die geplante Fragestellung nicht mit bereits bestehenden Datensätzen bearbeiten kann
- **interdisziplinäre Forschung =** Perspektive und Methodik unterschiedlicher Disziplinen auf den Datensatz

NACHTEILE BEIM TEILEN VON DATEN

- ggf. zusätzliche Arbeit bei
  - der Recherche nach geeigenten Repositorium
  - der Abklärung von Rechten Dritter
  - der Dokumentation der Daten und Materialien nötig
- Metadaten müssen erfasst werden
- Angst vor „Ideenklau“, Datenrestriktionen nur begrenzt und zeitlich beschränkt möglich
- Missbrauch ist theoretisch möglich, wenn das Repositorium nicht vertrauenswürdig ist

DATENGEBENDENBEFRAGUNG DES FDZ AM IQB (2018)

"Welche Hindernisse sehen Sie für die Publikation von Forschungsdaten?"



Mehrfachauswahl möglich, *N*=24

NACHTEILE BEIM NACHNUTZEN VON DATEN

- **keine sofortige Verfügbarkeit der neuesten Studien** = viele werden frühestens 1-2 Jahre nach Projektende/Berichtlegung an unser FDZ übergeben
- **teilweise mangelnde Dokumentation der Daten** = Rückfragen an Datenproduzierende oder FDZ notwendig
- **Passgenauigkeit zur eigenen Fragestellung nicht immer gegeben** = Variablen/Konstrukte fehlen oder sind nicht optimal operationalisiert
- **Restriktionen der Datenproduzierenden** betreffen intendierte Forschungsfrage

#Datenschatz heben? Jetzt bewerben!

Forschung braucht Neu-Gierde. Neu-Gierige brauchen Daten. Wir wissen von vielen neuen Daten, manchmal älter, manchmal ganz frisch. Doch oft nur schwer zu finden. Auf einer Projekthomepage oder in einem Repositorium – aber ohne die Metadaten oder die Dokumentation mit denen andere Neu-Gierige diese Daten finden könnten. Hier wird KonsortSWD helfen.

Ausschreibung Projektförderung Forschungsdatenmanagement

- Hier finden Sie die Ausschreibungsunterlagen →
- Bewerbungsende: 30. April 2021
- Bekanntgabe Förderung: Juni 2021
- Frühster Förderbeginn: 1. Juli 2021
- Weitere Ausschreibungsrunden u. a. 2024, 2025

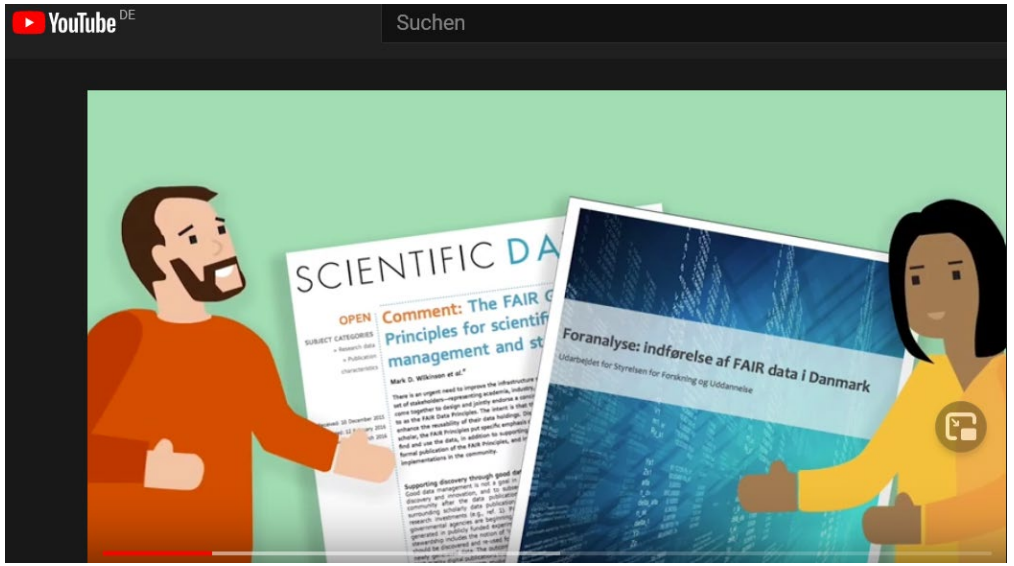
https://www.konsortswd.de/konsortswd/das-konsortium/services/

CR

DATEN TEILEN – WIE?

FAIR Prinzipien

- Findable
- Accessible
- Interoperable
- Re-usable



https://www.youtube.com/watch?v=5OeCrQE3HhE

https://www.go-fair.org/fair-principles/

FAIR Prinzipien

- Findable
- Accessible
- Interoperable
- Re-usable

Umsetzung ?

FAIR Aware

Data Archiving and Networked Services
DANS

Let's assume you have research data almost ready for uploading to a repository: do you already know how you and the repository can work together to make the data as findable, accessible, interoperable and reusable (FAIR) as possible? By guiding you through the assessment process, the FAIR-Aware tool can help you to better understand the FAIR Principles and how making data FAIR can increase the potential value and impact of your data.

FAIR-Aware is an disciplinary-agnostic online tool developed by the FAIRsFAIR project. Different scientific communities can adapt it to their own use. You should, however, have a target dataset in mind to be able to answer the questions and complete the assessment. The tool is not meant to give you a score for the FAIRness of a specific dataset.

This tool addresses data, but other research outputs can be made FAIR as well. If you are interested in what FAIR entails for research software, you may want to consult Five Recommendations for FAIR Software and Towards FAIR principles for research software.

The assessment starts with a few questions 'about you' followed by 10 questions about FAIR. For each question additional information and guidance can be displayed (click on the "i" icon). The majority of the questions will help you assess your current level of awareness about what actions are needed to make data FAIR. At the end, your feedback will help us improve FAIR-Aware and make it as user-friendly as possible. You will need between 10 and 30 minutes to complete the assessment depending on your familiarity with the subject and issues covered.

The FAIRsFAIR Team (DANS, DCC, UniHB)

Find out more about FAIRsFAIR on the project's website. If you have any questions, drop us an e-mail.

About you

Which research domain do you work in? Domain

Which of the following describes your role? Please select all that apply.

- ☐ Researcher
- ☐ Policy maker
- ☐ Research support (e.g. data steward, curator, data manager, librarian, information technology professional)
- ☐ Funder
- ☐ Publisher
- ☐ Other

Which of the following types of organisations best describe your employer? Please select all that apply.

- ☐ Research Infrastructure/eInfrastructure (e.g. data repository, service provider, library)
- ☐ University
- ☐ Research Performing Organisation
- ☐ Government
- ☐ eInfrastructure (e.g. repository or scientific data provider)
- ☐ Funding Body
- ☐ Publisher
- ☐ Industry
- ☐ Other

https://fairaware.dans.knaw.nl/

HOW TO FAIR

What is FAIR | Why FAIR | How to FAIR | About | Quiz

How to make your data FAIR

This page will show you how you can make your research data more FAIR by taking you through six FAIRification practices:

1. Documentation
2. File formats
3. Metadata
4. Access to data
5. Persistent identifiers
6. Data licences

https://howtofair.dk/how-to-fair/

Thüringer Kompetenznetzwerk Forschungsdatenmanagement

Wettbewerb „FAIRest Dataset“ - Teilnahmebedingungen

1. **Veranstalter des Wettbewerbs**
   Technische Universität Ilmenau (TU Ilmenau)
   Der Rektor
   Ehrenbergstraße 29
   98693 Ilmenau

2. **Teilnahmefrist**
   Die Teilnahme am Wettbewerb ist möglich vom 15. Februar bis zum 30. April 2021.

3. **Gegenstand des Wettbewerbs**
   Gegenstand des Wettbewerbs sind veröffentlichte Datensätze, die sich an den FAIR-Prinzipien[1] orientieren. Es gewinnt der Datensatz, der die FAIR-Prinzipien am besten erfüllt. Das bedeutet die Daten sollen auffindbar, zugänglich, interoperabel und wiederverwendbar sein.

4. **Vergabe und Verwendung des Preisgelds**
   Das Preisgeld beträgt bis zu 2000,00 Euro. Nach Ablauf der Teilnahmefrist erfolgt die Auswertung der eingesendeten Beiträge durch die fünf Mitglieder des Thüringer Kompetenznetzwerks Forschungsdatenmanagement unter Zuhilfenahme einschlägiger *self assessment tools*. Im Falle eines Gewinnes werden die Teilnehmer/innen vom Kompetenznetzwerk Forschungsdatenmanagement per E-Mail über den Gewinn informiert. Wenn sich die Gewinner/innen nicht innerhalb von vier Wochen auf die Gewinnbenachrichtigung hin gemeldet haben, verlieren sie den Gewinnanspruch.
   Das Preisgeld kann nur an Einrichtungen oder Projekte an einer Thüringer Hochschule vergeben werden. Die Auszahlung und Abwicklung erfolgt dabei durch die Finanzverwaltung der jeweiligen Hochschule.
   Das Preisgeld muss nachweisbar für Forschungsdatenmanagement verwendet werden. Das bedeutet, dass hiermit Personal oder Sachkosten finanziert werden können, die im

https://forschungsdaten-thueringen.de/nachricht/fairest-dataset-award-2021-de.html

Datendokumentation: Was gehört alles dazu?

Informationen zur Datenaufbereitung

- Transkriptionsregeln, Codierregeln
- Kodierungen, Gewichtungsvariablen, Imputierte Werte, sonstige generierte Variablen
- Angaben zu durchgeführten Plausibilitäts-, Konsistenz- und Fehlerkontrollen

Die Instrumente, die zur Messung eingesetzt wurden

- Fragebögen
- Listen- und Kartensätze
- Testinstrumente
- Leitfäden
- Beobachtungsprotokolle
- Ratingbögen

Informationen über die Messung (den Prozess), d.h. die Feldphase (u.a. Paradaten)

- Intervieweranweisungen
- Kontaktprotokolle
- Notizen zum Kontext, Setting (wo und wie fand die Messung statt, wer war anwesend...)

Informationen über den Hintergrund der Messung/Datenerhebung (Metadaten)

- Methoden-/ Feldberichte
- Zwischen- und Abschlussberichte
- Publikationen der Forschungsergebnisse bspw. in wissenschaftlichen Fachzeitschriften
- Websites

Informationen zu Datenschutz und Anonymisierung

- an die Probanden ausgehändigte Einverständniserklärungen oder Informationsblätter zum Datenschutz
- Vorgehen bei der Anonymisierung (Anonymisierungsprotokolle)
- Liste über Pseudonyme
- Einschätzung des Re-Identifikationsrisikos, Nennung sensibler Informationen in den Daten

Informationen zur Datenanalyse, zum Vorgehen der Datenanalyse

- QDA-Dateien
- Syntax

https://www.forschungsdaten-bildung.de/welchedaten?la=de

LEITFRAGEN

DOKUMENTATION VON DATEN

- Welche Informationen benötigen nicht am Projekt beteiligte Dritte, um eine Sekundärnutzung der Forschungsdaten durchführen zu können?
- Welche Informationen benötige ich, um in unbestimmter Zukunft die durchgeführten Analysen replizieren zu können?
- Welche Informationen benötigen Projektmitarbeiter, die nicht an der Datenerhebung beteiligt waren?
- Welche Informationen benötigen Projektmitarbeiter, die nicht an der Datenaufbereitung und/oder den Datenanalysen mitgearbeitet haben?

forschungsdaten bildung

https://www.forschungsdaten-bildung.de/dokumentieren

Checkliste Daten vorbereiten für Übergabe

Variablen- und Wertebenennung	☐ Variablen und Werte sind nach einem konsistenten Schema verständlich und eindeutig bezeichnet (z. B. V1 bis V100). ☐ Alle Variablen sind über ihren Namen und/oder ihre Labels den jeweiligen Items im Fragebogen bzw. den Konstrukten im Skalenhandbuch zuordenbar. ☐ Die Namen sind möglichst kurz gewählt. ☐ Sonderzeichen, Umlaute oder Leerzeichen wurden nicht verwendet.
Variablen- und Wertelabels	☐ Idealerweise: Variablenlabels wurden für alle Variablen verwendet. ☐ Idealerweise: Alle Werte sind gelabelt. ☐ Variablen- und Wertelabels sind eine kurze und möglichst aussagekräftige Beschreibung der Variableninhalte bzw. -ausprägungen. ☐ Variablen- und Wertelabels enthalten keine Sonderzeichen, Umlaute und Leerzeichen.

https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-4_v1_2.pdf

Checkliste Daten vorbereiten für Übergabe

Variablenwerte	☐ Alle Variablenausprägungen sind jeweils einem eindeutigen numerischen Wert zugewiesen. ☐ Die zugewiesenen numerischen Werte folgen, soweit möglich, einem einheitlichen Schema. ☐ Offene Antwortmöglichkeiten sind ggf. codiert, d. h. mit einem numerischen Wert versehen worden. ☐ Offene Antwortmöglichkeiten sind auf datenschutzrechtliche Probleme untersucht worden.
Fehlende Werte	☐ Fehlende Werte sind definiert. ☐ Fehlende Werte sind als solche durch ein Label gekennzeichnet. ☐ Existieren verschiedene Arten von fehlenden Werten, sind diese möglichst differenziert festgehalten.

https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-4_v1_2.pdf

Checkliste Daten vorbereiten für Übergabe

Plausibilitäts- und Konsistenz-prüfungen	☐ Die Filterführung ist korrekt. ☐ Die Variablenausprägungen sind plausibel. ☐ Es sind keine wild codes (Werte außerhalb des zulässigen Wertebereichs) vorhanden.
Dateinamen	☐ Dateinamen enthalten keine Sonderzeichen, Umlaute oder Leerzeichen. ☐ Dateinamen sind möglichst kurz gewählt. ☐ Dateien sind so bezeichnet, dass aus dem Namen Rückschlüsse auf Studie, Daten-/Materialtyp und ggf. Versionsnummer gezogen werden können.
Datenformat	☐ Die Dateien liegen in einem empfohlenen oder zumindest akzeptablen Format vor.

https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-4_v1_2.pdf

Checkliste Daten vorbereiten für Übergabe

Dokumentations-materialien

- ☐ Instrumente der Datenerhebung, bspw. Fragebögen, werden eingereicht.
- ☐ Ein Codebuch oder Skalenhandbuch enthält alle im Datensatz auftretenden Variablen.
- ☐ Zu jeder Variable ist deutlich, welcher Frage im Fragebogen sie entspricht und wie die Antwortalternativen kodiert wurden.
- ☐ Falls Variablen erstellt wurden (z. B. Skalenbildung, Index, abgeleitete Variablen) ist dies im Code-/Skalenhandbuch nachvollziehbar beschrieben.
- ☐ Angaben zum Untersuchungsdesign und der Feldphase sind in einem Methodenhandbuch/Feldbericht festgehalten.

https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-4_v1_2.pdf

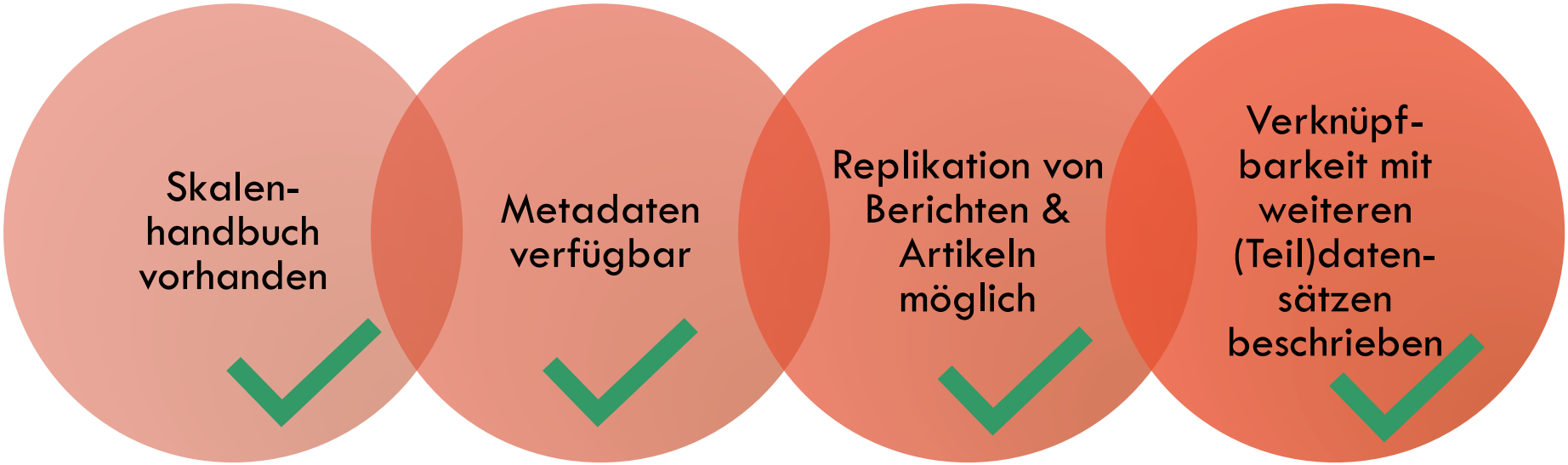
Checkliste Daten vorbereiten für Übergabe

Datenschutz und rechtliche Aspekte

- ☐ Eigennamen von Personen wurden gelöscht und durch nicht sprechende Identifikatoren ersetzt, z. B. id1 bis id198 (formale Anonymisierung).
- ☐ Weitere personenbezogene Daten (u. a. Eigennamen von Orten und Organisationen) liegen nicht vor oder die Studie wurde anonymisiert (formale Anonymisierung).
- ☐ Regionale, berufliche und ähnliche Angaben sind nicht so kleinteilig, dass die Informationen die Identifizierung der Teilnehmer/-innen ermöglichen (andernfalls sind Hinweise an das Archiv erfolgt).
- ☐ Es liegen keine urheberrechtlichen oder vertraglichen Hindernisse vor, die einer Archivierung oder Datenweitergabe im Wege stehen.

https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-4_v1_2.pdf

ANFORDERUNGEN AN DIE DOKUMENTATION DER DATEN AM FDZ AM IQB



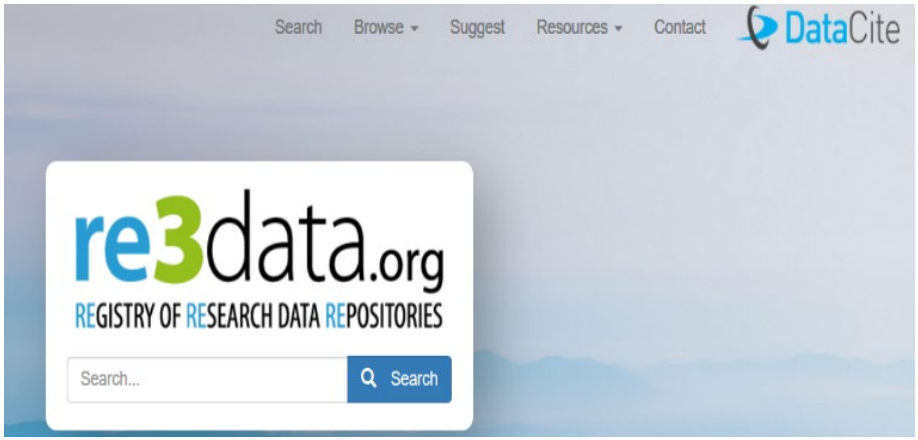
DATEN TEILEN – WO?

DATEN TEILEN – WO?

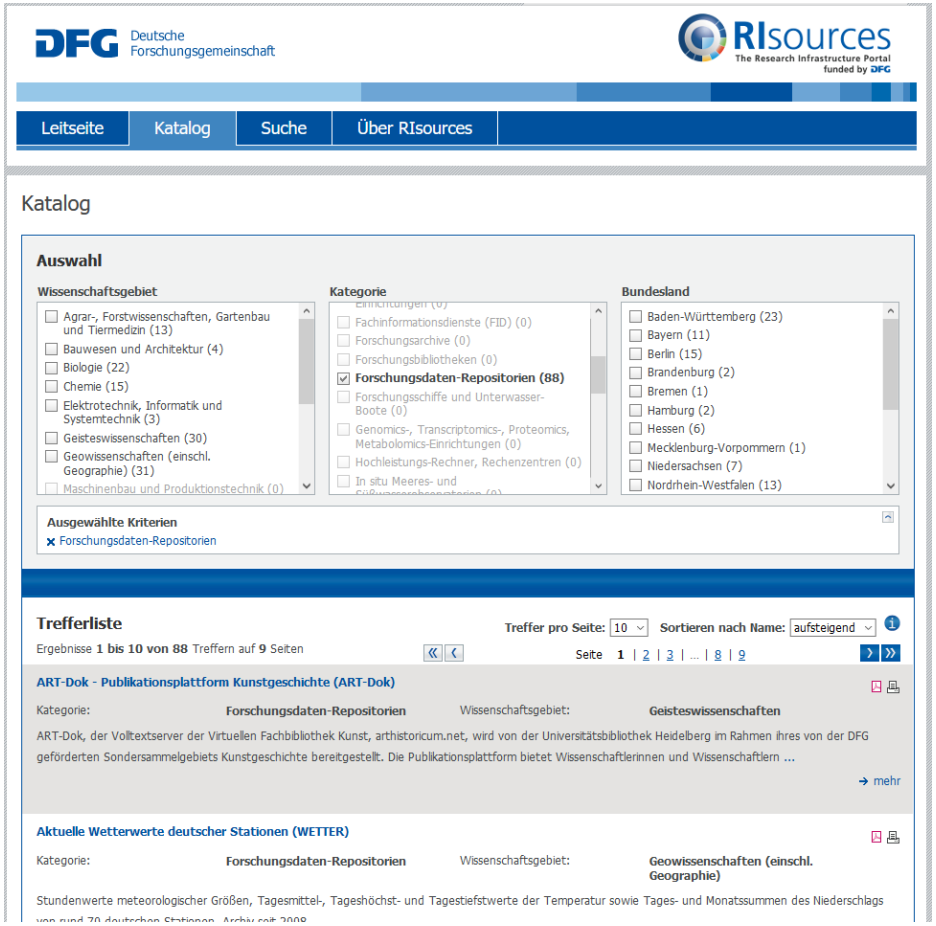
- Forschungsdatenzentren
- Universitätsbibliothekenund Universitäten
- Zeitschriften
- Daten-Zeitschriften
- Plattformen

LEITFRAGEN DATEN-REPOSITORIEN

- Ist es zertifiziert?
- Werden persistente Identifier vergeben?
- Wie ist der Datenzugang geregelt?
- Gibt es Nutzungsbedingungen/-beschränkungen?



https://www.re3data.org/



https://risources.dfg.de/

REPOSITORIEN-SUCHE

Forschungsdatenzentren

Repositorium
Forschungsdatenzentrum des Deutschen Zentrum für Altersfragen
Forschungsdatenzentrum am Robert Koch Institut
Nationales Bildungspanel
Forschungsdatenzentrum des SOEP
fdz.DZHW
Forschungsdatenzentrum Bildung
eLabour
ZEW Forschungsdatenzentrum
Forschungsdatenzentrum Internationale Umfrageprogramme
Bundeszentrale für Gesundheitliche Aufklärung Studien
Forschungsdatenzentrum im Bundesinstitut für Berufsbildung
Forschungsdaten- und Servicezentrum der Bundesbank
Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt und Berufsforschung
Forschungsdatenzentrum Wahlen
Forschungsdatenzentrum Wissenschaftsstatistik
International Data Service Center of the Institute for the Study of Labor
Forschungsdatenzentrum der Rentenversicherung
Forschungsdatenzentrum PIAAC bei GESIS
Forschungsdatenzentrum IWH
Forschungsdatenzentrum im Kraftfahrt-Bundesamt
PsychData
Datenbank Gesprochenes Deutsch
Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder
FDZ German Microdata Lab
FDZ-BO
Forschungsdatenzentrum des Deutschen Jugendinstituts
Survey of Health, Ageing and Retirement in Europe
Forschungsdatenzentrum Ruhr am RWI
ALLBUS
Pairfam
Qualiservice
Verbund Forschungsdaten Bildung
International Service of Geomagnetic Indices
The World Data Center for Remote Sensing of the Atmosphere
International Earth Rotation and Reference Systems Service
World Data Centre for Geomagnetism, Copenhagen
World Stress Map
Biobank of blood donors
CDC - Climate Data Center
DSMZ

Liste zertifizierter Forschungsdaten-repositorien in Deutschland

https://edoc.hu-berlin.de/handle/18452/22762

Disziplinspezifische Forschungsdatenzentren

https://www.psychdata.de/

https://www.forschungsdaten-bildung.de/daten-uebermitteln?la=de

Disziplinspezifische Forschungsdatenzentren

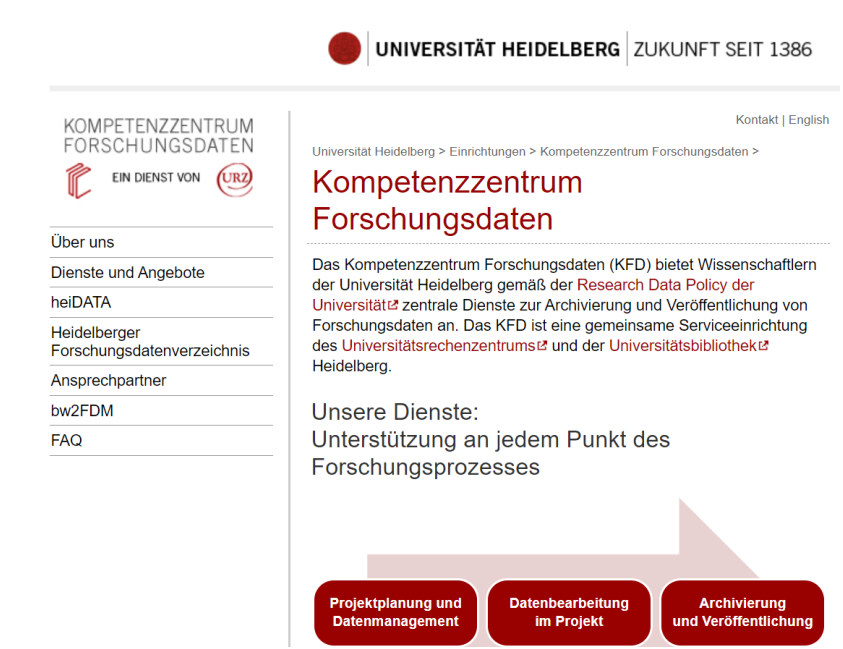
KonsortSWD
Konsortium für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften

RatSWD.
Rat für Sozial- und WirtschaftsDaten

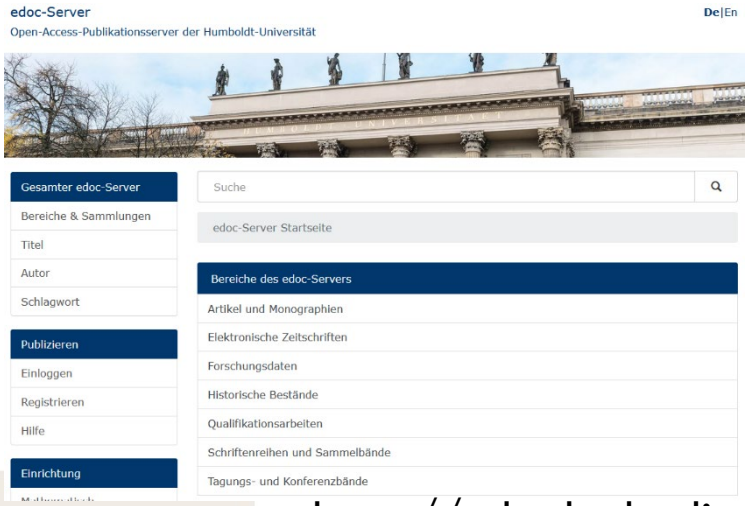
https://www.konsortswd.de/datenzentren/alle-datenzentren/

AKKREDITIERTE DATENZENTREN DES RATSWD

Universitätsbibliotheken und Universitäten



https://data.uni-heidelberg.de/



https://edoc.hu-berlin.de/

Direkt zu GRO.data

Betrieben durch:
GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

https://data.goettingen-research-online.de/

DATEN IN ZEITSCHRIFTEN-REPOSITORIEN VERÖFFENTLICHEN?

Zeitschriften

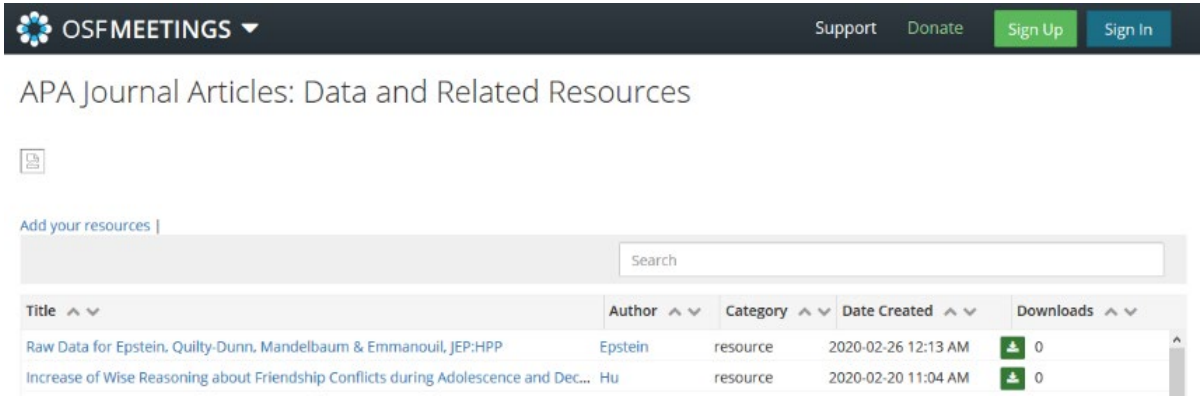
Vorteile:

- Artikel & Daten sind an der gleichen Stelle und gut auffindbar
- Verlage haben eine hohe Reichweite
- Artikel und dazugehörige Daten sind gut auffindbar (Daten allein aber nicht)

Nachteile:

- Verlage folgen meist nicht den FAIR-Prinzipien:
- Ggf. ist Datenabgabe bei Verlagen mit Rechteabgabe verbunden
- Es ist nicht gesichert, dass Daten bei Verlagen langfristig archiviert und dauerhaft verfügbar sind

APA kooperiert mit dem Open Science Framework als Data Repository



Data Journals

Hier wird eine Liste von Data Journals gepflegt, die vorwiegend Data Papers publiziere
Ergänzungen sind erwünscht.

- Atomic Data and Nuclear Data Tables (Elsevier)
- Biodiversity Data Journal (Pensoft Publishers)
- Biomedical Data Journal (Procon Ltd.)
- BMC Research Notes (Biomed Central)
- Chemical Data Collections (Elsevier)
- Data (MDPI)
- Data in Brief - Zusammenschluss u.a. 2017 mit Genomics Data (Elsevier)
- Data Observer (DeGruyter)
- Dataset Papers in Science - *2017 eingestellt* (Hindawi Publishing Corporation)
- Earth System Science Data - ESSD (Copernicus Publications)
- Ecological Archives (Ecological Society of America - ESA/Wiley)
- F1000Research (F1000 Research)
- Geoscience Data Journal (Wiley)
- Geoscientific Model Development (Copernicus)
- GigaScience (Oxford University Press)
- Internet Archaeology (Internet Archaeology)
- International Journal of Epidemiology (Oxford University Press)
- The Astrophysics Journal: Supplement Series (American Astronomial Society)
- The International Journal of Robotics Research (Sage Publications)
- Journal of Open Archaeology Data (JOAD) (Ubiquity Press)
- Journal of Open Health Data (Ubiquity Press)
- Journal of Open Psychology Data (JOPD) (Ubiquity Press)
- Journal of Open Research Software (JORS)(Ubiquity Press)
- Journal of Chemical & Engineering Data (ACS Publications)
- Journal of Physical and Chemical Reference Data (AIP Publishing)
- Nuclear Data Sheets (Elsevier)
- Open Data Journal for Agricultural Research (diverse)
- Open Journal of Bioresources (Ubiquity Press)
- Research Data Journal for the Humanities and Social Sciences (Brill)
- RIDE (A review journal for digital editions and resources)
- Scientific Data (Nature Publishing Group)

https://www.forschungsdaten.org/index.php/Data_Journals

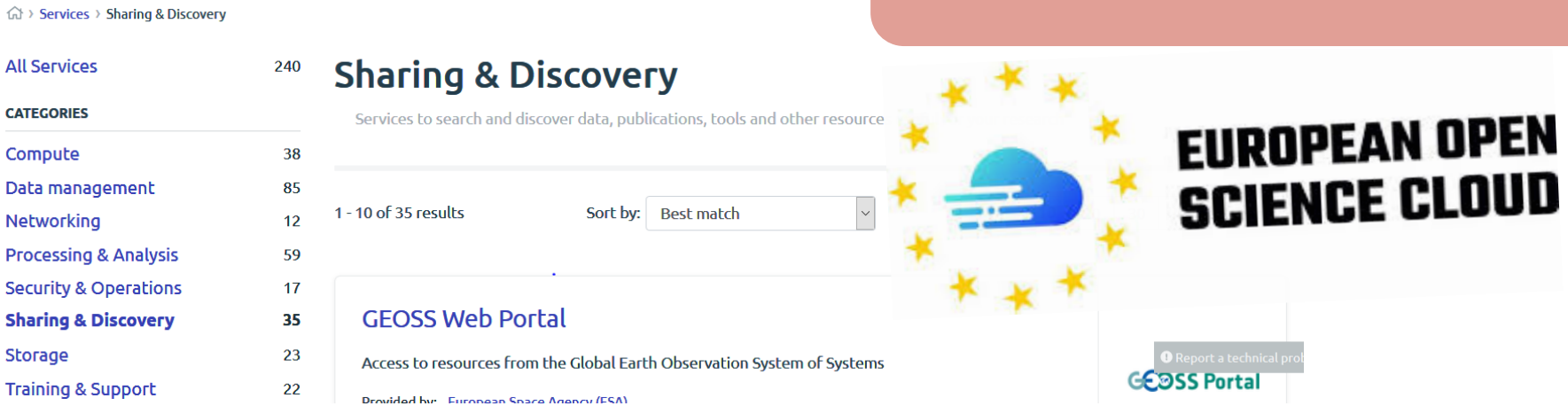
Daten-Zeitschriften

https://openpsychologydata.metajnl.com/

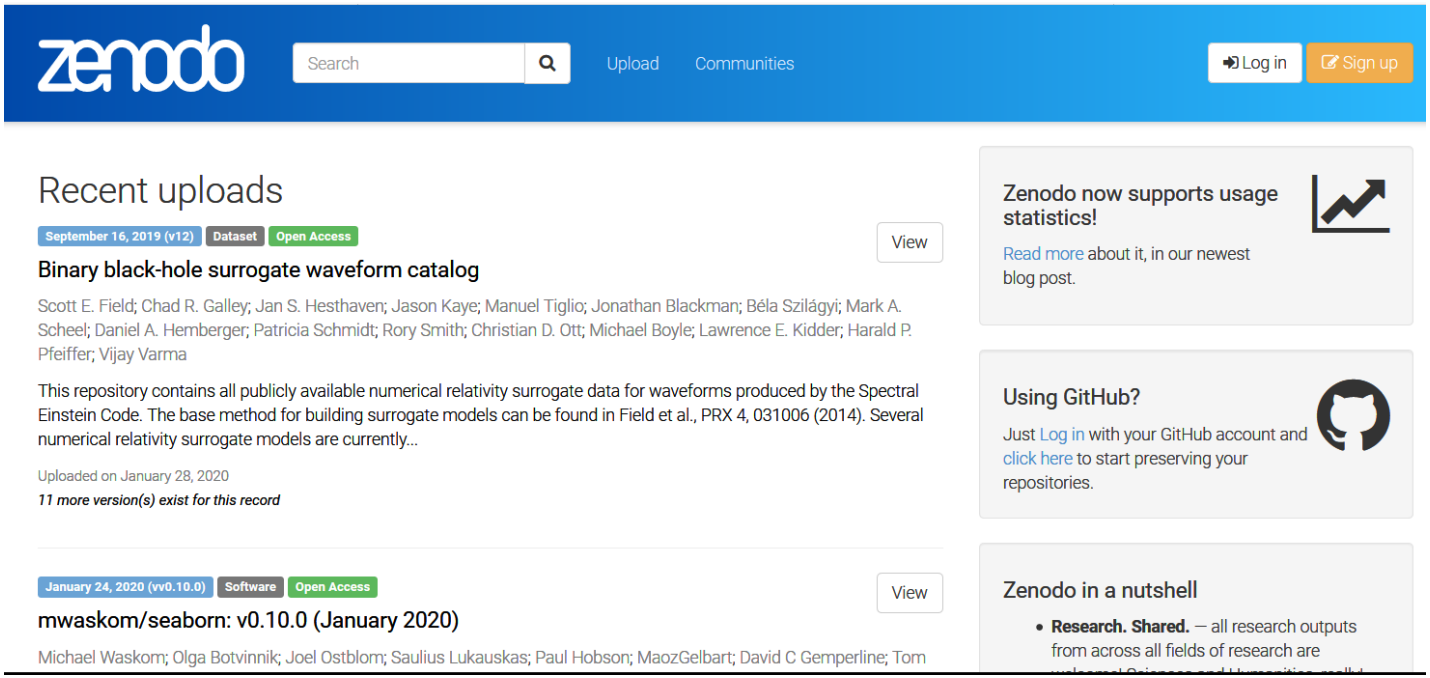
Plattformen

https://osf.io/

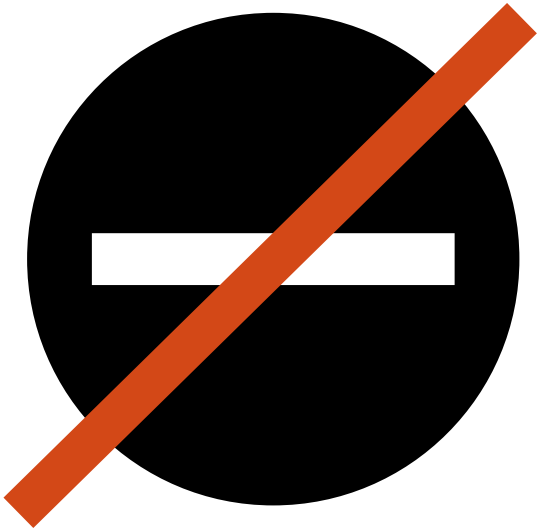
Plattformen



https://marketplace.eosc-portal.eu/

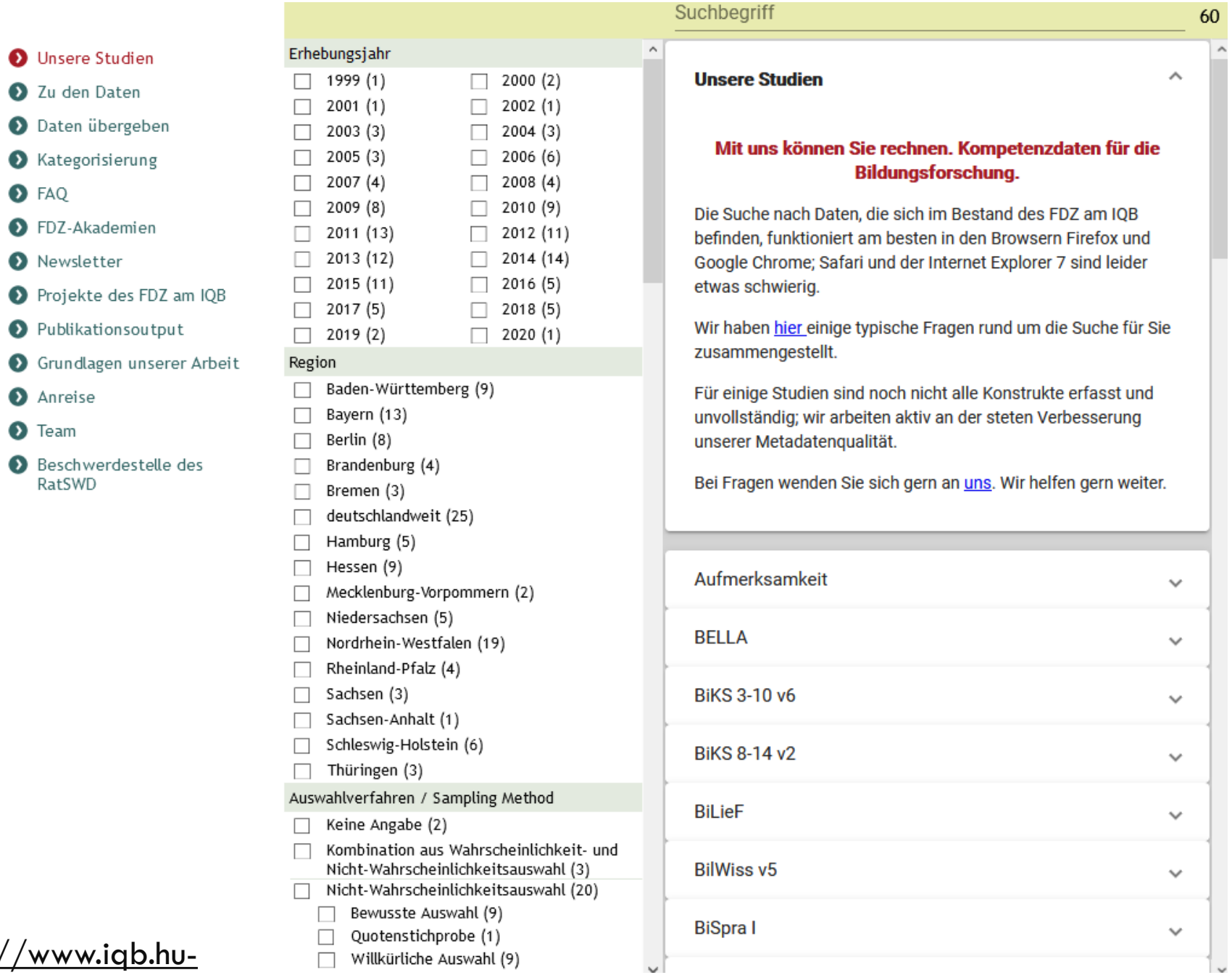


https://zenodo.org/

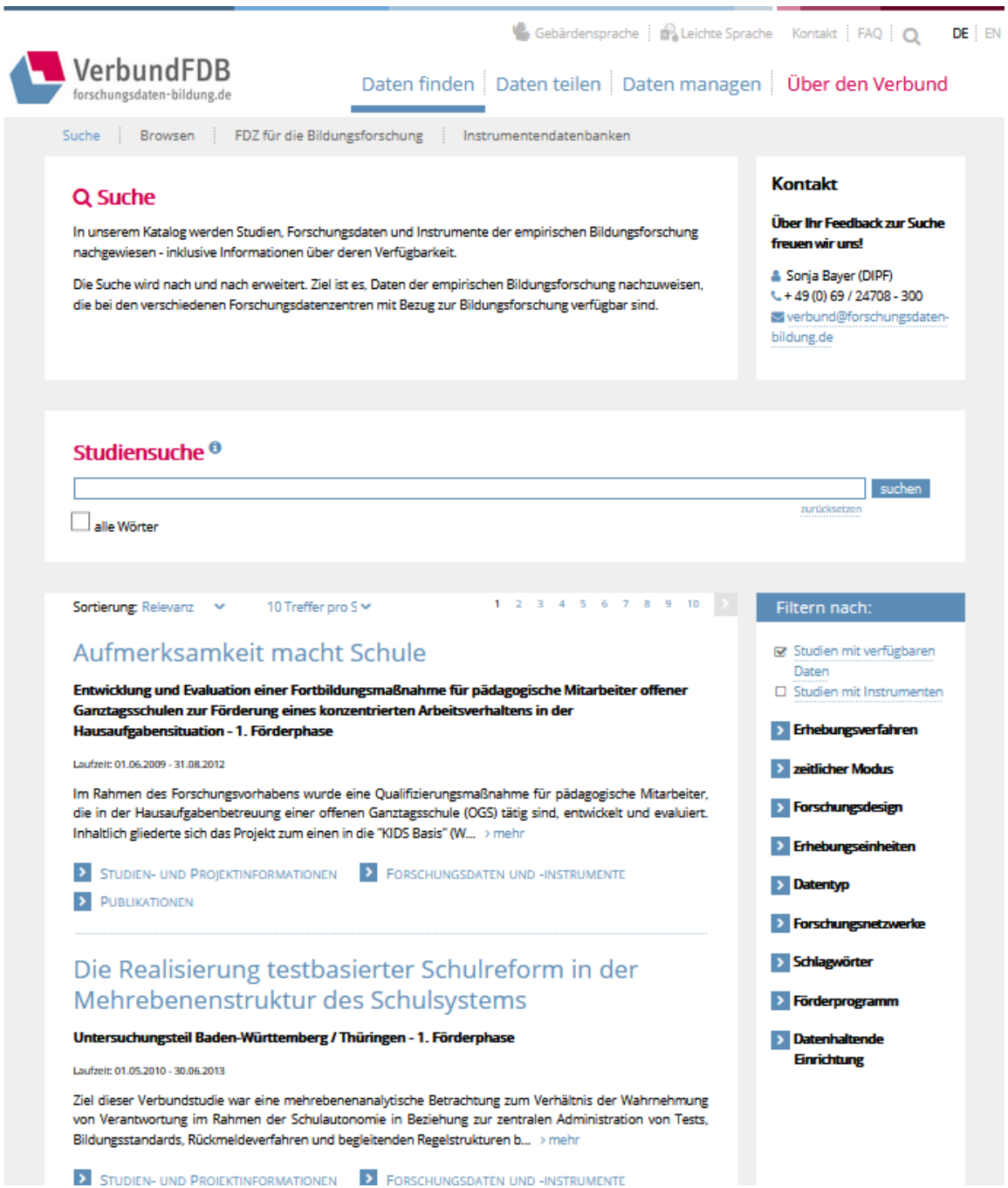


DATENSÄTZE FINDEN – WO?

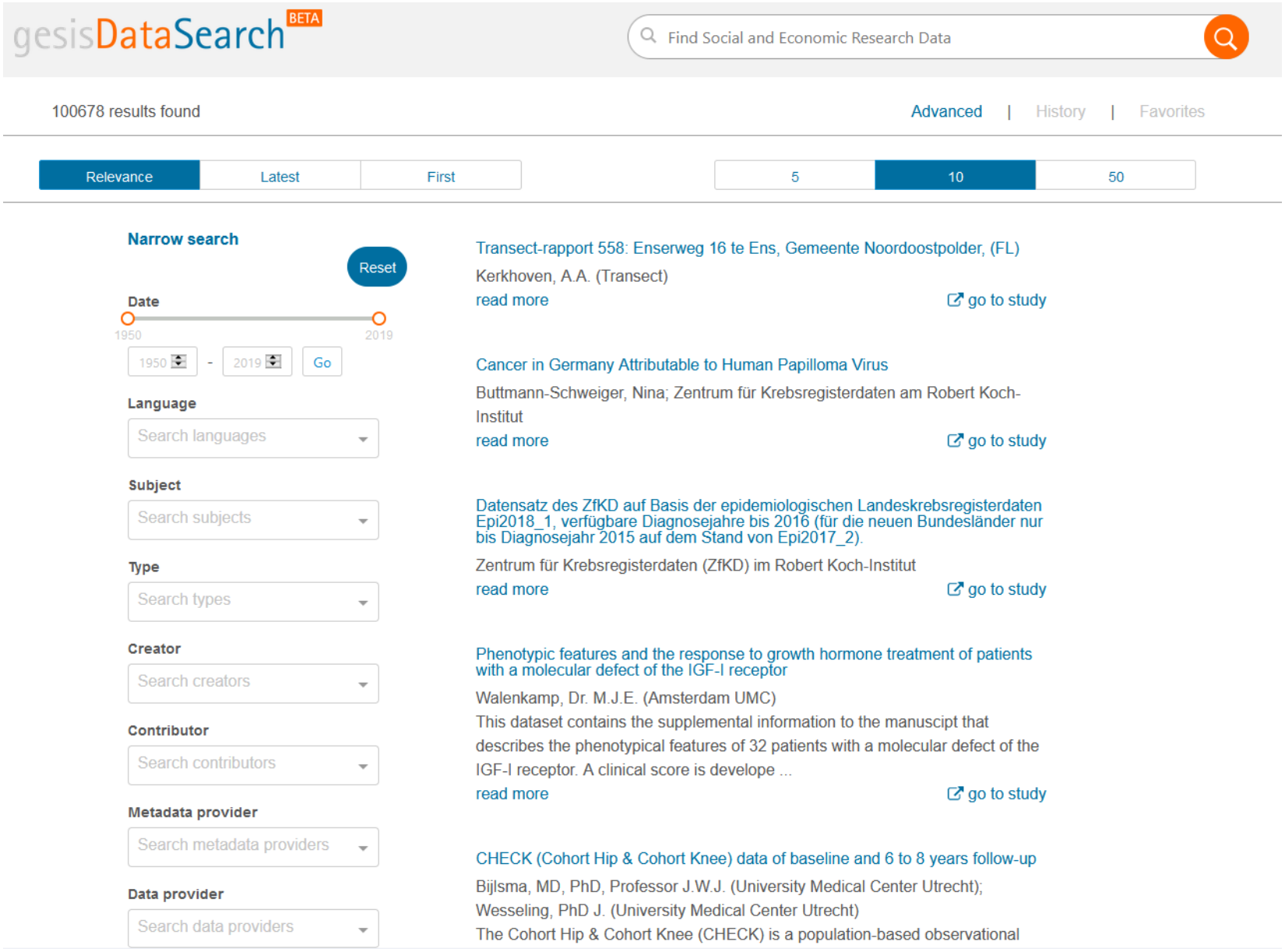
https://datasetsearch.research.google.com/



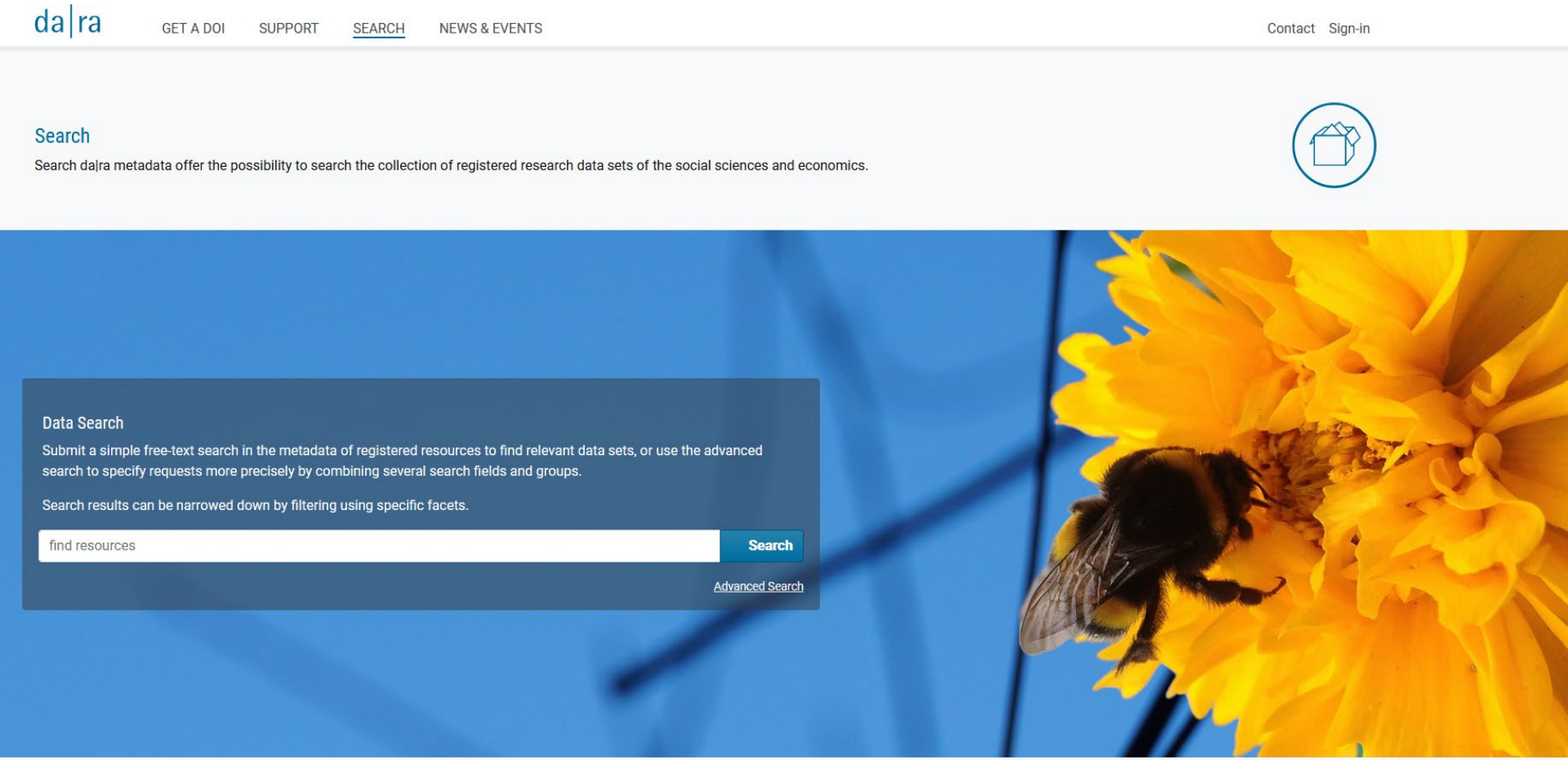
https://www.iqb.hu-berlin.de/fdz/studies/



https://www.forschungsdaten-bildung.de/studienliste.php?la=de



https://datasearch.gesis.org/search



https://www.da-ra.de/search

https://www.konsortswd.de/datenzentren/datensuche-in-den-fdz/

DataCite Commons

NEPS

Pages Support Sign In

Works People Organizations

407 Works

Publication Year

- 2021 6
- 2020 71
- 2019 68
- 2018 51
- 2017 67
- 2016 48
- 2015 26
- 2014 27
- 2013 17
- 2012 15
- 2011 5
- 2010 1

Work Type

- Text 259
- Dataset 91
- Physical Object 18
- Collection 12
- Image 4
- Software 4
- Audiovisual 2
- Other 2

License

- CC-BY-4.0 20
- CC-BY-SA-4.0 5
- CC-BY-NC-3.0 1
- CC0-1.0 1

Language

- German 110
- English 81
- Catalan 1

NEPS-Startkohorte 6: Erwachsene (SC6 1.0.0)

Hans-Peter Blossfeld
Version 1.0.0 of Content published 2011 in LIfBi Leibniz Institute for Educational Trajectories

Other Identifiers
DOI: https://doi.org/10.5157/neps:sc6:1.0.0

DOI registered December 8, 2011 via DataCite.

Dataset

https://doi.org/10.5157/neps:sc6:1.0.0

Signal intensity distributions of probes measuring annotated exons (EPs), non-exon regions (NEPs) and negative controls (NCPs)

Gabor Halasz, Marinus F Van Batenburg, Joelle Perusse, Sujun Hua, Xiang-Jun Lu, Kevin P White & Harmen J Bussemaker
Figure published 2011 in figshare Academic Research System

Copyright information: Taken from "Detecting transcriptionally active regions using genomic tiling arrays"Genome Biology 2006;7(7):R59-R59.Published online 19 Jul 2006PMCID:PMC1779562. Cumulative distribution functions of signal intensities for these probe populations are shown for a representative channel.

DOI registered June 8, 2012 via DataCite.

Image

https://doi.org/10.6084/m9.figshare.45591

NEPS-Startkohorte 2: Kindergarten (SC2 1.0.0)

Hans-Günther Roßbach & NEPS, Nationales Bildungspanel/National Educational Panel Study, Bamberg (Germany)
Version 1.0.0 of Content published 2012 in LIfBi Leibniz Institute for Educational Trajectories

DOI registered September 14, 2012 via DataCite.

Dataset German

https://doi.org/10.5157/neps:sc2:1.0.0

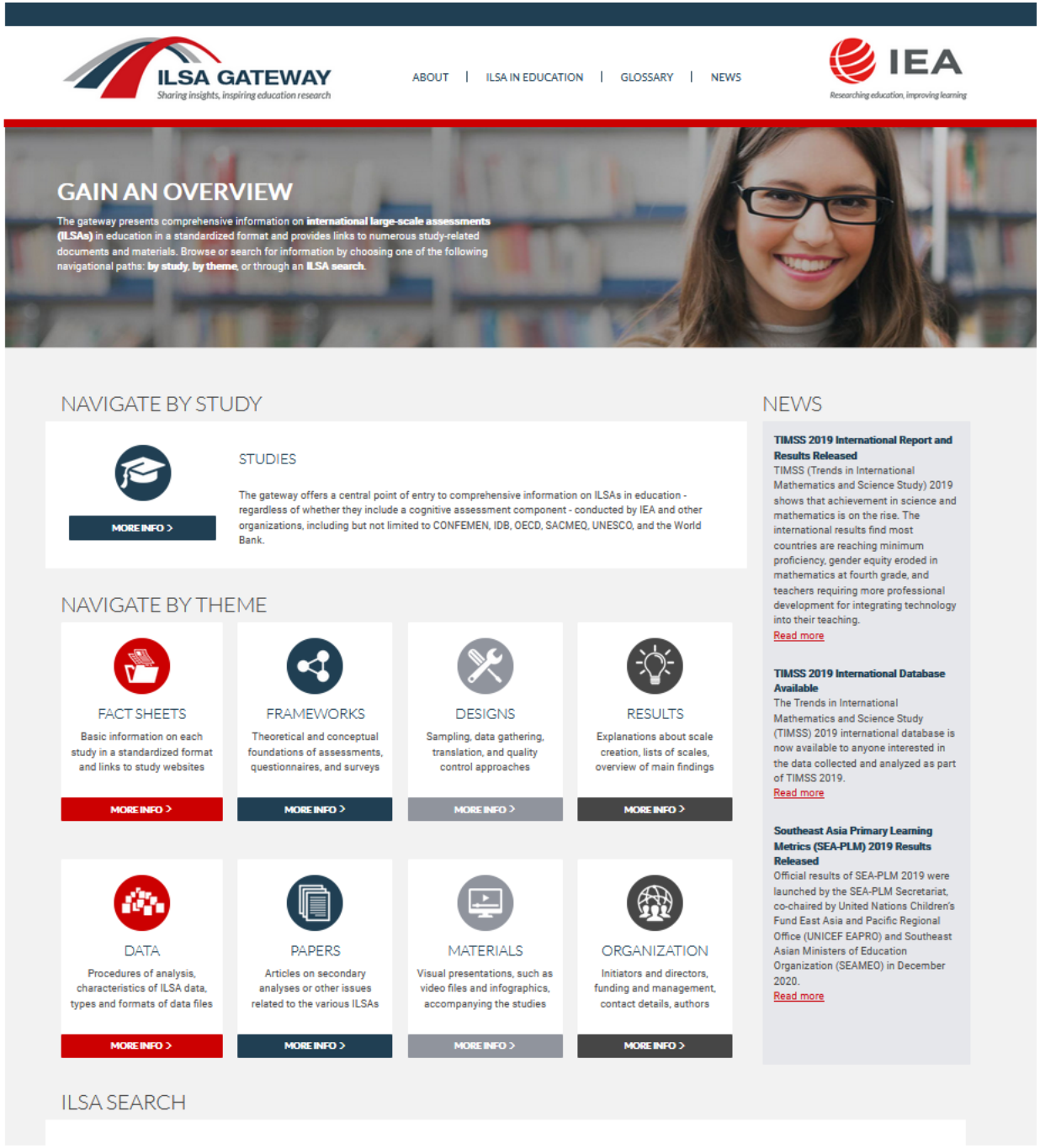
NEPS-Startkohorte 3: Klasse 5 (SC3 1.0.0)

Hans-Günther Roßbach & NEPS, National Educational Panel Study, Bamberg (Germany)
Version 1.0.0 of Content published 2012 in LIfBi Leibniz Institute for Educational Trajectories

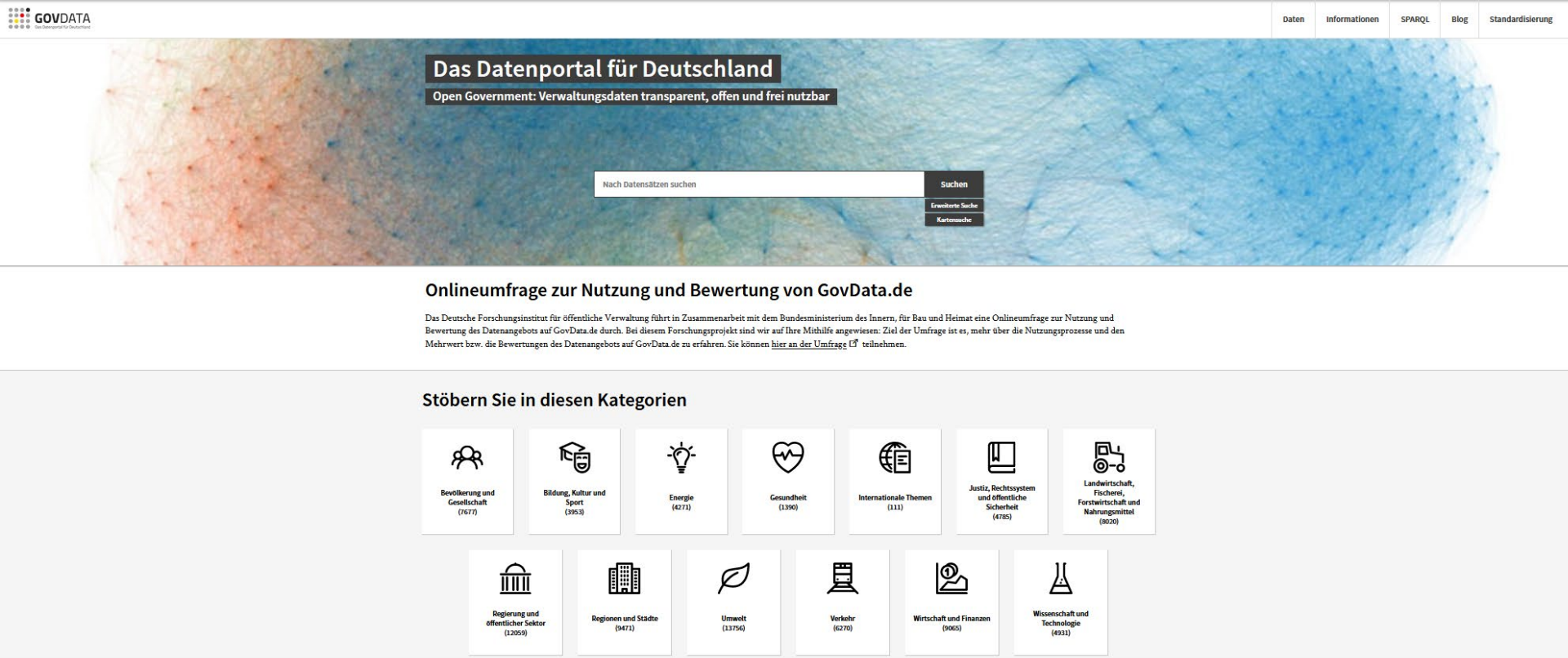
DOI registered September 14, 2012 via DataCite.

Dataset German

https://datacite.org/



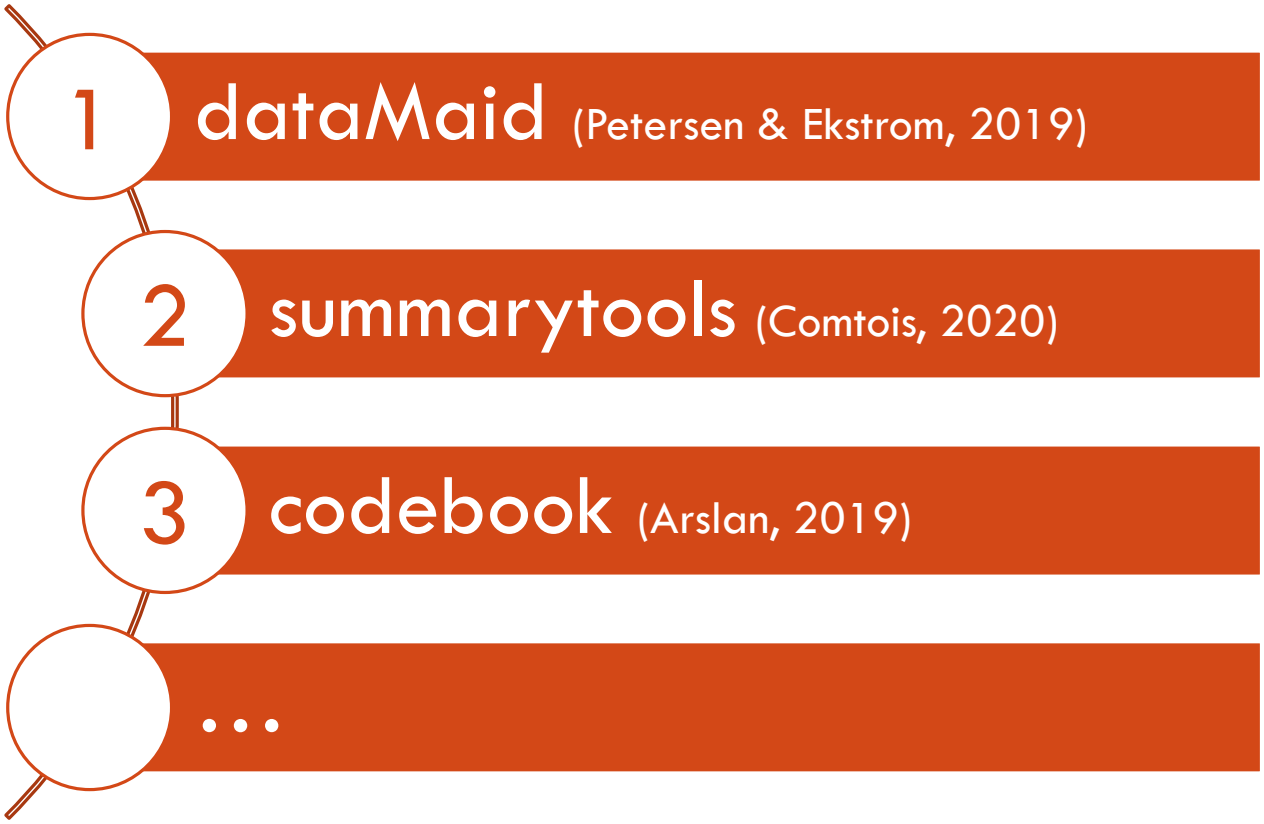
https://www.ilsa-gateway.org/



https://www.govdata.de/

https://www.bmwi.de/Redaktion/DE/Artikel/Service/Gesetzesvorhaben/zweites-open-data-gesetz-und-datennutzungsgesetz.html

Skalenhandbuch in R erstellen: 3 geeignete Pakete



- es gibt zahlreiche Pakete, die die Datendokumentation erleichtern und automatisieren
- diese Pakete eignen sich auch gut, um sich einen Überblick über den Datensatz zu verschaffen

Skalenhandbuch in R erstellen: Vorbereitung – fortgeschrittenes Datenmanagement

- erster Schritt: fortgeschrittenes Datenmanagement
  - Variablen- und Wertelabel ergänzen
  - Skalenbildung
  - Analyse fehlender Werte

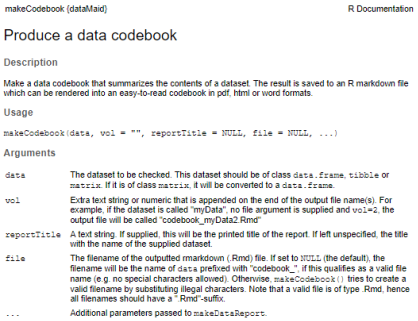
- mit verschiedenen R-Paketen können diese Informationen herangespielt werden (z.B. direkt in R, über eine Excel-Tabelle, aus einem SPSS-Datensatz etc.)
  - misty (Yanagida, 2020): Datenmanagement (auch gut geeignet für Mehrebenenstrukturen)
  - eatGADS(Becker, 2021): Umstrukturierung von Datensätzen, Einlesen von SPSS-Datensätzen mit Metadaten, Datenbankformate → besonders gut geeignet für Large-Scale Daten (mit Imputationen und Standardfehler-Anpassungen
  - psych (Revelle, 2020): Skalenbildung etc.
  - …

BREAKOUT SESSIONS

3 Gruppen – Skalenhandbuch mit…

(1) „dataMaid“

```
install.packages("dataMaid")
library(dataMaid)
?dataMaid::makeCodebook
```



makeCodebook (dataMaid) R Documentation

Produce a data codebook

Description

Make a data codebook that summarizes the contents of a dataset. The result is saved to an R markdown file which can be rendered into an easy-to-read codebook in pdf, html or word formats.

Usage

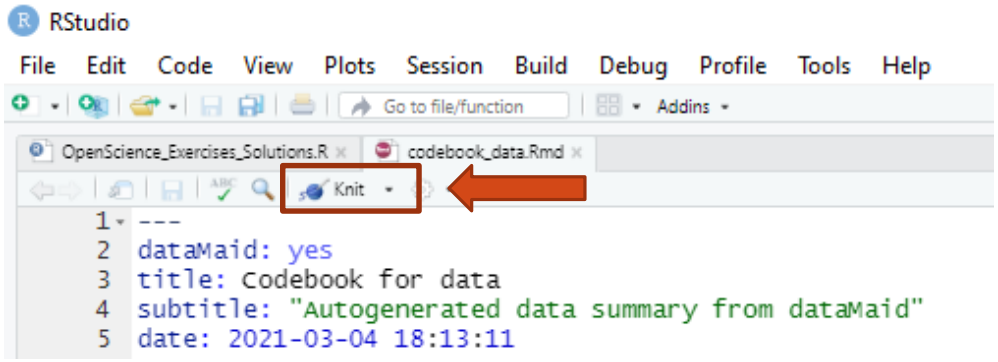
```
makeCodebook(data, vol = "", reportTitle = NULL, file = NULL, ...)
```

Arguments

data	The dataset to be checked. This dataset should be of class data.frame, tibble or matrix. If it is of class matrix, it will be converted to a data.frame.
vol	Extra text string or numeric that is appended on the end of the output file name(s). For example, if the dataset is called "myData", no file argument is supplied and vol=2, the output file will be called "codebook_myData2.Rmd"
reportTitle	A text string. If supplied, this will be the printed title of the report. If left unspecified, the title with the name of the supplied dataset.
file	The filename of the outputted rmarkdown (.Rmd) file. If set to NULL (the default), the filename will be the name of data prefixed with "codebook_". If this qualifies as a valid file name (e.g. no special characters allowed). Otherwise, makeCodebook() tries to create a valid filename by substituting illegal characters. Note that a valid file is of type .Rmd, hence all filenames should have a ".Rmd"-suffix.
...	Additional parameters passed to makeDataReport.

Achtung: erzeugte .rmd-Datei muss ggf. noch mal als UTF-8 abgespeichert werden:

1. rmd-Datei in RStudio öffnen
2. File→Save with Encoding…→ UTF-8→ OK
3. anschließend: “Knit”-Button klicken:



AK

BREAKOUT SESSIONS

3 Gruppen – Skalenhandbuch mit…

(2) „summarytools“

```
install.packages("summarytools")
library(summarytools)
?summarytools::dfSummary
view(dfSummary(data))
```

Data frame Summary

Description

Summary of a data frame consisting of: variable names, labels if any, factor levels, frequencies and/or numerical summary statistics, and valid/missing observation counts.

Usage

```
dfSummary(
  x,
  round.digits = 1,
  varnumbers = st_options("dfSummary.varnumbers"),
  labels.col = st_options("dfSummary.labels.col"),
  valid.col = st_options("dfSummary.valid.col"),
  na.col = st_options("dfSummary.na.col"),
  graph.col = st_options("dfSummary.graph.col"),
  graph.magnif = st_options("dfSummary.graph.magnif"),
  style = st_options("dfSummary.style"),
  plain.ascii = st_options("plain.ascii"),
  justify = "l",
  col.widths = NA,
  headings = st_options("headings"),
  display.labels = st_options("display.labels"),
  max.distinct.values = 10,
  trim.strings = FALSE,
  max.string.width = 25,
  split.cells = 40,
  split.tables = Inf,
  tmp.img.dir = st_options("tmp.img.dir"),
  silent = st_options("dfSummary.silent")
```

AK

BREAKOUT SESSIONS

3 Gruppen – Skalenhandbuch mit…

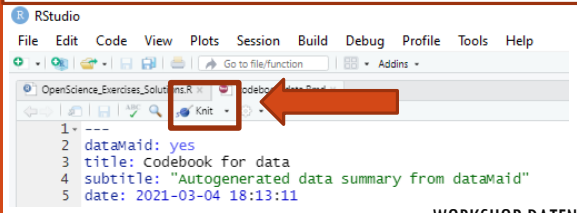
(3) „codebook“

```
install.packages("codebook")
library(codebook)
```

Tutorial:
https://rubenarslan.github.io/codebook/articles/codebook_tutorial.html

Skalenhandbuch als html-Dokument erstellen:

1) RMarkdown Dokument öffnen: skalenhandbuch_mit_codebook_template.Rmd
2) Pfad mit gespeichertem Datensatz anpassen (Zeile 74)
3) wenn nötig: Name des Datensatzes anpassen (Zeile 76ff.;in unserem Beispiel heißt der Datensatz "data")
4) wenn nötig weitere Anpassungen im Text und den Metadaten vornehmen
5) Blauen "Knit"-button klicken, um Skalenhandbuch zu erzeugen



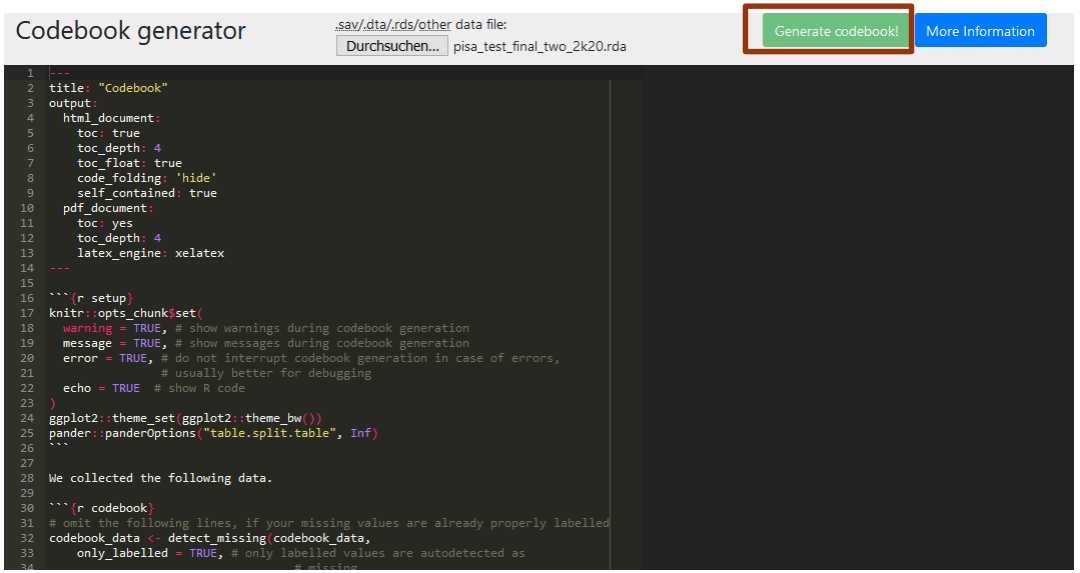
AK

BREAKOUT SESSIONS

3 Gruppen – Skalenhandbuch mit…

(3) „codebook“

Alternative: Skalenhandbuch online erstellen: https://codebook.formr.org



DATENSCHUTZ & SENSIBLE VARIABLEN

The Strava Heat Map and the End of Secrets

The US military is reexamining security policies after fitness tracker data shared on social media revealed bases and patrol routes

Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Latest: Strava suggests military users 'opt out' of heatmap as row deepens

The Guardian, 28.01.2018:
https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases

WIRED, 29.01.2018:
https://www.wired.com/story/strava-heat-map-military-bases-fitness-trackers-privacy/

DATENSCHUTZ & SENSIBLE VARIABLEN

Svea Eckert (2017, in *The Guardian*):

„What would you think if somebody showed up at your door saying: 'Hey, I have your complete browsing history – every day, every hour, every minute, every click you did on the web for the last month'? How would you think we got it: some shady hacker? No. It was much easier: you can just buy it."

The Guardian, 01.08.2017: https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers

DATENSCHUTZ & SENSIBLE VARIABLEN

6 Apr 14

Datenskandal an österreichischen Schulen, BIFIE-Direktor wird gekündigt

Wie veröffentlicht wurde, sind Daten der "Informellen Kompetenzmessung" (IKM) zur Bildungsstandarderhebung aus den Jahren 2011 und 2012 sowie E-Mail-Adressen von 37.000 österreichischen Lehrern auf einem rumänischen Server aufgetaucht. Die Daten waren dort bei einem Subunternehmer nicht verschlüsselt abgelegt worden und somit öffentlich zugänglich.

https://www.itexperst.at/datenskandal-oesterreichischen-schulen-bifie-direktor-wird-gekuendigt-7023.html

Ausnahmeregelung mit OECD

Die von Bildungsministerin Gabriele Heinisch-Hosek (SPÖ) verordnete Auszeit beim PISA-Test für 2015 ist nun zurückgenommen worden. Die „Kronen Zeitung" berichtete vorab von Verhandlungen mit der OECD, die nun eine Ausnahmeregelung ermöglichen.

https://orf.at/v2/stories/2231204/2231207/

Abo · E-Paper · Magazine

WIENER ZEITUNG .at

MENÜ | POLITIK | KULTUR | WIRTSCHAFT | AMTSBLATT | DOSSIERS | MEINUNG

Startseite › Politik › Österreich

SCHÜLERDATENSKANDAL

Skandal oder doch nicht

- **Wegen Datenaffäre Rücktritt von Unterrichtsministerin Heinisch-Hosek gefordert**
- **Ex-Bifie-Direktor Haider: "Die Daten sind weitgehend wertlos."**

vom 27.02.2014, 17:25 Uhr | Update: 27.02.2014, 21:50 Uhr

https://www.wienerzeitung.at/nachrichten/politik/oesterreich/611778-Skandal-oder-doch-nicht.html

DER STANDARD › Inland

BILDUNGSTESTS

Bifie-Datenleck gestopft, Tests gestoppt

Alle Daten der Zentralmatura und der Bildungstests sind laut STANDARD-Infos in einer "Hochsicherheitsumgebung" gespeichert

Lisa Nimmervoll 26. Februar 2014, 17:29 168 Postings

Wien - Angesichts des "größten Datenskandals der österreichischen Schulgeschichte" forderten die Lehrergewerkschaften am Mittwoch den

https://www.derstandard.at/story/1392686483213/bifie-datenleck-gestopft-tests-gestoppt

DATENSCHUTZ & SENSIBLE VARIABLEN

Auf der einen Seite…

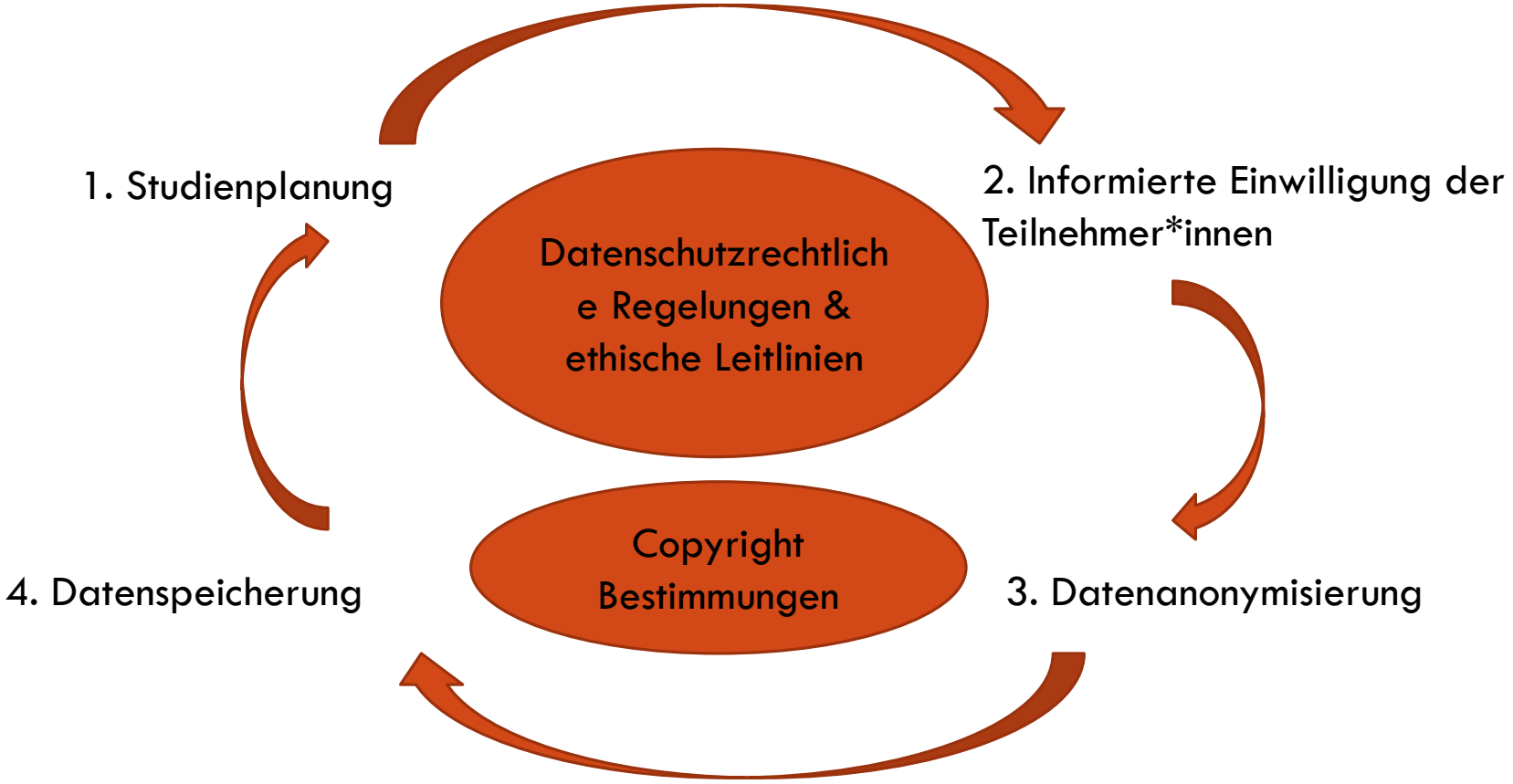
…verlangen immer mehr Zeitschriften und Mittelgeber, der Scientific Community die Forschungsdaten (in anonymisierter Form) bereitzustellen

Auf der anderen Seite…

…müssen dabei datenschutzrechtliche Bestimmungen eingehalten werden und Forscher*innen können nicht sicher sein, ob einzelne Personen in den Daten identifizierbar sind

DATENSCHUTZ & SENSIBLE VARIABLEN

Ethische und datenschutzrechtliche Fragen müssen während des gesamten Forschungsprozesses berücksichtigt werden



Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

Meyermann, A. & Porzelt, M. (2019, S.6). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

Ethische Richtlinien

4 Prinzipien (DGPs, 2018)

- Respekts vor Selbstbestimmung (Freiwilligkeit, Teilnehmerinformationen, Verständlichkeit, Einwilligung)
- Nichtschädigung
- Fürsorge
- Gerechtigkeit

Literaturtipp: Meyer (2018) - Practical Tips for Ethical Data Sharing

- Liste mit DOs und DON'Ts

Invited Forum: Challenges in Making Data Available
Tutorial

Practical Tips for Ethical Data Sharing

Michelle N. Meyer
Geisinger Health System, Danville, Pennsylvania

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(1) 131–144
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2515245917747656
www.psychologicalscience.org/AMPPS
SAGE

DGPs (2018). Ethisches Handeln in der psychologischen Forschung. Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen. Göttingen: Hogrefe

datenschutzrechtlich relevante Variablen

a) direkte Identifier
= personenbezogen

- Vor- und Nachnamen
- Adressen
- Geburtstage
- Versicherungsnummern
- Telefonnummern
- Email-Adressen
- …

b) indirekte Identifier
= personenbeziehbar

- Vornamen
- Stadt
- Straße
- Schulnamen
- Postleitzahlen
- Berufsangaben
- Bilder
- Datumsangaben
- Sprache zu Hause
- Geburtsort
- Seltene Merkmale (n≤5)

c) besondere Kategorien personenbezogener Daten

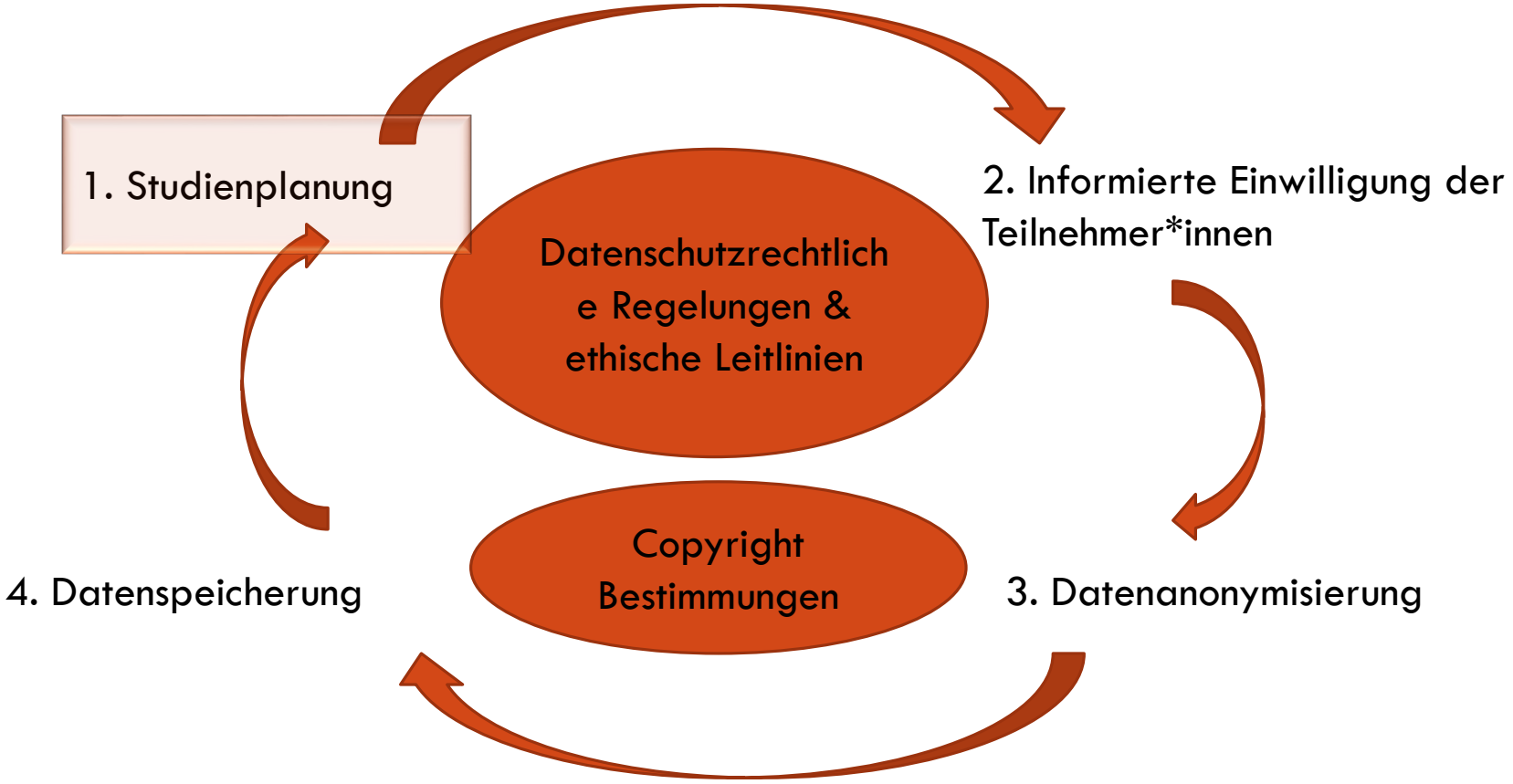
- Ethnizität
- Politische Einstellungen
- Religion
- Sexuelle Orientierung
- Medizinische und genetische Angaben

→ Erfassung dieser Merkmale muss in Einwilligungserklärung beschrieben werden

Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

Ethische und datenschutzrechtliche Fragen müssen während des gesamten Forschungsprozesses berücksichtigt werden



Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

1. Studienplanung – Prinzip der Datensparsamkeit & Verhältnismäßigkeit

- nur die personenbezogenen Daten erheben, die zur Beantwortung der Forschungsfragen erforderlich sind
- angemessenen Detailgrad bei der Erfassung und größtmöglichen Anonymisierungsgrad wählen
- Zugang zu erhobenen personenbezogenen Daten sollte so wenigen Personen im Forschungsteam wie möglich gewährt werden
- schon vor Datenerhebung sollte Datenmanagementplan erstellt werden: Umgang mit sensiblen Daten festlegen und abstimmen

→ Datenschutzbeauftragte der Universität/ des Forschungsinstituts helfen und beraten

Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

1. Studienplanung – Leitfragen

- An welcher Stelle im Forschungsprojekt werden personenbezogene Daten verarbeitet und in welchem Umfang?
- Wie kann deren Verarbeitung so geregelt werden, dass die Eingriffe für die Betroffenen möglichst gering sind?
- Könnte auf die Erhebung und Verarbeitung personenbezogener Daten verzichtet werden?
- Welcher Personenkreis hat zu welchen Zeitpunkten Zugriff auf die Daten?
- Müssen und können Einwilligungen zur Verarbeitung der Daten eingeholt werden?
- Können die Daten anonymisiert werden? Wie schnell und wie umfassend?
- Wie können die Daten geschützt aufbewahrt werden?

Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

Ethische und datenschutzrechtliche Fragen müssen während des gesamten Forschungsprozesses berücksichtigt werden

Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

2. Informierte Einwilligung der Teilnehmer*innen einholen

Abbildung 3: Bestandteile von Einwilligungserklärungen

Zentrale Bestandteile von Einwilligungserklärungen

- Angabe des Forschungsvorhabens (Titel, Ziele, Forschungsfragen)
- Ablauf des Vorhabens (z. B. „Es werden Interviews geführt"; „Unterricht wird per Video aufgezeichnet")
- Nennung eines Ansprechpartners, Nennung des Verantwortlichen für die Datenerhebung und -verarbeitung
- Nennung des zuständigen Datenschutzbeauftragten
- Hinweise auf die Rechte des oder der Betroffenen (Freiwilligkeit, Widerrufsrecht, Folgenlosigkeit bei Verweigerung oder Widerruf)
- Angabe der Daten, die erhoben werden (z. B. Kompetenzen, Einstellungen, Verhalten, Adressdaten)
- Angabe Art der Datenerhebung (z. B. per Fragebogen, per Video)
- Angabe der geplanten Verarbeitung der Daten (z. B. Digitalisierung, Transkription, Kodierung, Anonymisierung)
- Angabe der Verwendungszwecke, d. h. Angabe der geplanten Datennutzungen (z. B. Auswertung, Veröffentlichung, Nutzung in der Lehre, Archivierung, Weitergabe der Daten für Sekundärnutzungen)

Meyermann, A. & Porzelt, M. (2019, S. 14). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

2. Informierte Einwilligung der Teilnehmer*innen einholen

Fallbeispiel 1: Angabe unterschiedlicher Verwendungszwecke im Einwilligungsteil

Simplifiziertes Beispiel für eine abgestufte Abfrage	Simplifiziertes Beispiel für eine gebündelte Abfrage
Ich bin einverstanden ...mit der Teilnahme an der Studie und der Verwendung meiner personenbezogenen Daten. ☐ Ja / ☐ nein ...mit der weiteren Nutzung meiner Daten – über dieses Projekt hinaus – für die Bildungsforschung allgemein. ☐ Ja / ☐ nein Datum, Unterschrift	Mit den voranstehend (im Informationsteil) gemachten Ausführungen zur Verwendung meiner personenbezogenen Daten bin ich einverstanden. Datum, Unterschrift

Meyermann, A. & Porzelt, M. (2019, S. 16). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

2. Informierte Einwilligung der Teilnehmer*innen einholen

enge vs. weite Zweckbindung

Enge Zweckbindung: Angabe eines Personenkreises und eines konkreten Zwecks;
z. B.: „Daten werden nur im Rahmen des Projektes zur Untersuchung der Einflüsse von Lehrerhandeln auf … genutzt und nur durch die genannten Projektmitarbeiter/innen bearbeitet.“;
Nachteil: Etwaige Zweckänderungen bedürfen einer erneuten Einwilligung.

Weite Zweckbindung: Angabe eines allgemeinen Verwendungszwecks und Nennung eines (unbestimmten) Personenkreises;
z. B. „Die Daten werden ausschließlich zu Zwecken der Bildungsforschung durch ausgewiesene Wissenschaftler/innen verwendet.“;
Vorteil: Eine Nachnutzung der Daten bspw. über Forschungsdatenzentren ist möglich. Nachteil: Unbekannte Effekte auf die Teilnahmebereitschaft.

- Checkliste: https://www.forschungsdaten-bildung.de/files/fdbinfo_1.pdf
- Beispiele: https://www.forschungsdaten-bildung.de/files/fdbinfo_4.pdf

Meyermann, A. & Porzelt, M. (2019, S. 19-20). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

Ethische und datenschutzrechtliche Fragen müssen während des gesamten Forschungsprozesses berücksichtigt werden

Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

Lösungsansätze zur Einhaltung datenschutzrechtlicher Bestimmungen

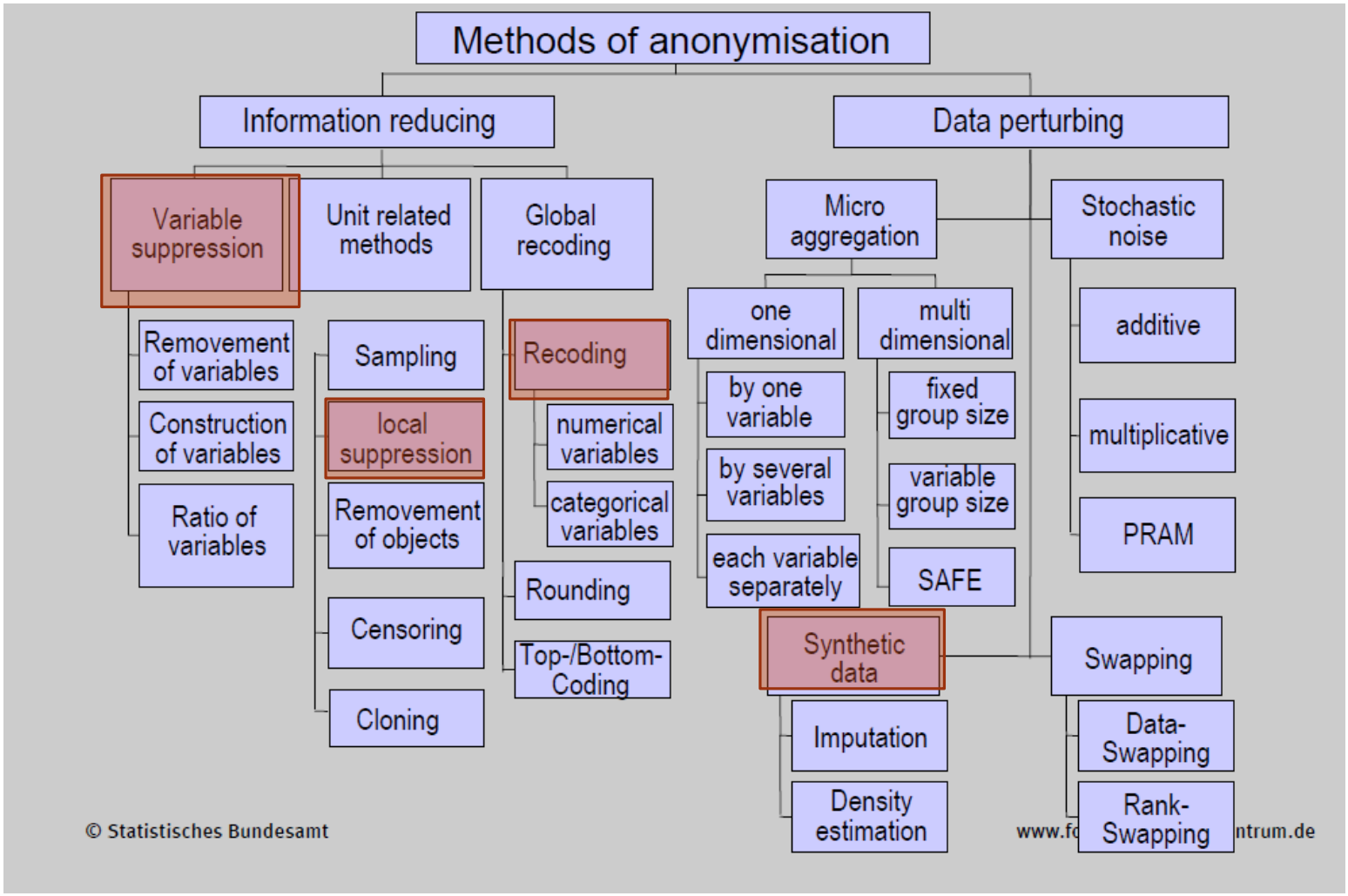


Abbildung vom Statistischen Bundesamt (www.forschungsdatenzentrum.de)

DATENSCHUTZ & SENSIBLE VARIABLEN

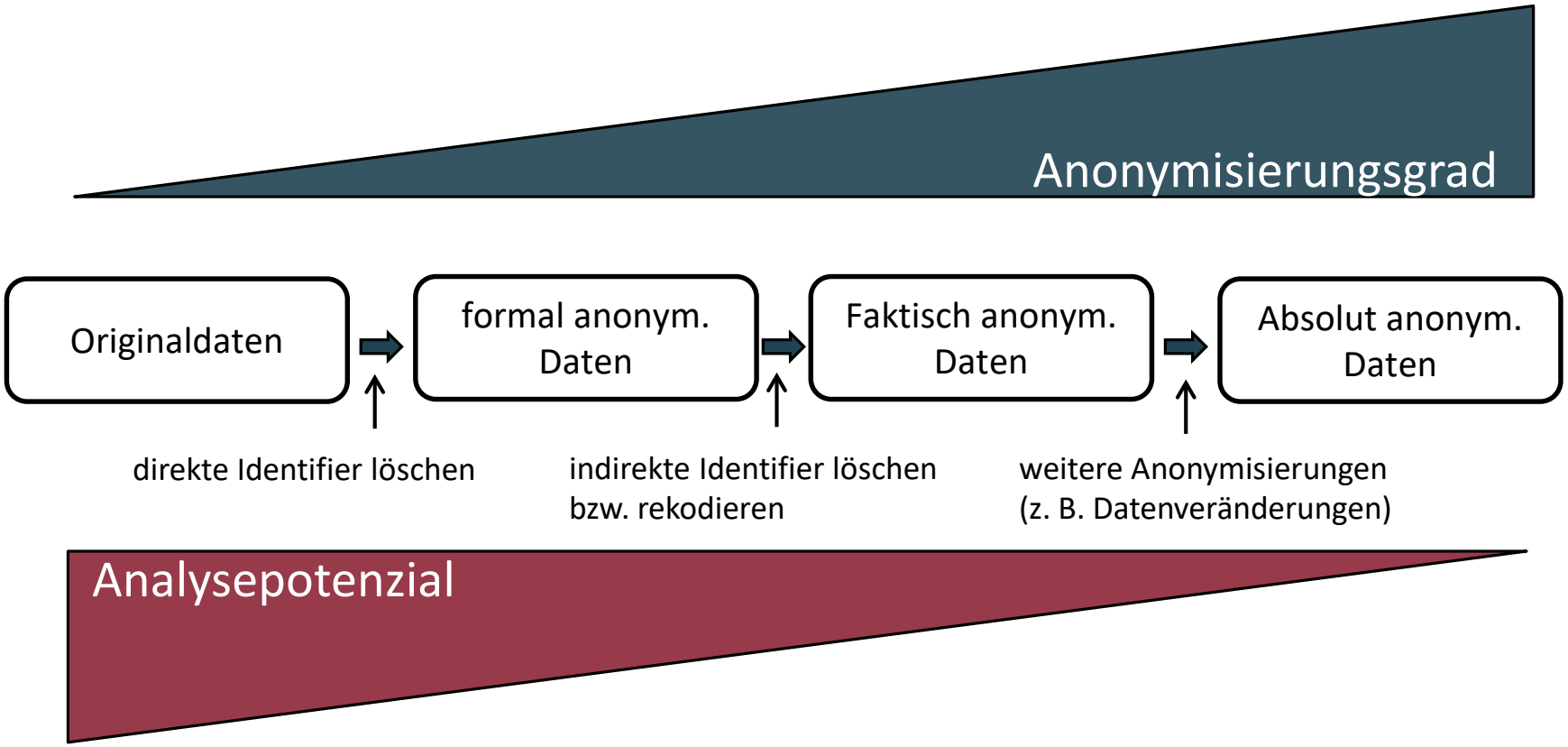
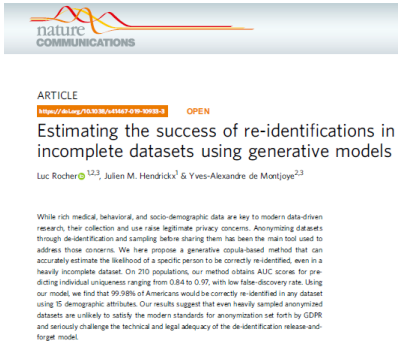
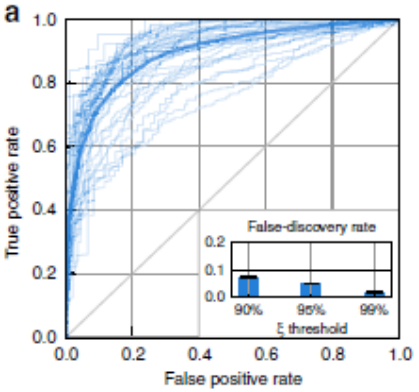


Abbildung vom Statistischen Bundesamt (www.forschungsdatenzentrum.de)

DATENSCHUTZ & SENSIBLE VARIABLEN

- Studie von Rocher et al. (2019):
  - höhere Anforderungen durch die EU-Datenschutzgrundverordnung; der Begriff faktische Anonymität ist nicht mehr relevant
  - faktische Anonymität = Merkmalsträger können nur mit sehr hohem Aufwand identifiziert werden (der in keinem Verhältnis zum Aufwand steht)
  - klassische Strategien zum Umgang mit datenschutzrechtlich bedenklichen Variablen sind im Internet- und Datenzeitalter nicht mehr so wirksam
  - Rocher et al. konnten zeigen, dass Kombination von Hintergrundmerkmalen (Geschlecht, Ethnizität, Geburtstag, Postleitzahl) ausreicht, um einzelne Personen eindeutig zu identifizieren
  - Methode: generatives graphisches Modell, welches mit sehr kleinen Teildatensätzen trainiert wurde (1%-Sample aus dem Gesamtdatensatz)

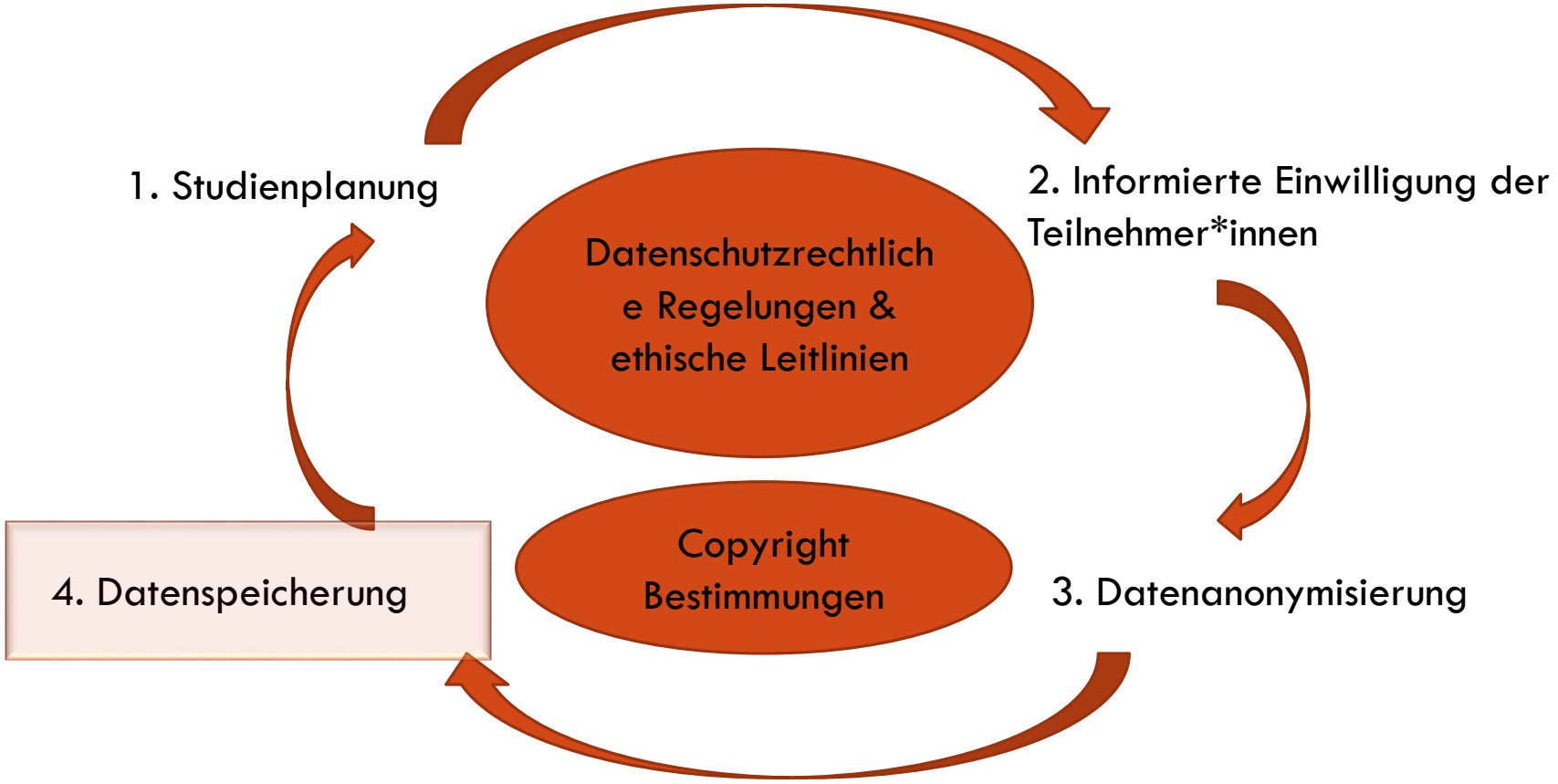




Rocher et al. (2019): https://doi.org/10.1038/s41467-019-10933-3

DATENSCHUTZ & SENSIBLE VARIABLEN

Ethische und datenschutzrechtliche Fragen müssen während des gesamten Forschungsprozesses berücksichtigt werden



Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

DATENSCHUTZ & SENSIBLE VARIABLEN

4. Datenspeicherung

- Ziel: personenbezogene Daten vor unbefugten Zugriffen, Manipulation und Verlust zu schützen
- hier helfen Datenschutzbeauftragt und das Rechenzentrum der Universität gut weiter
- Leitlinien zum Umgang mit Forschungsdaten (z. B. DFG, 2020; DGfE, GEBF & GFD, 2020) anwenden
- zur Dateienübermittlung universitäre Dateienaustauschdienste nutzen
- Datenzugang über gestaffelte Zugangswege (z. B. an Forschungsdatenzentren)

Tabelle 1: Beispiele für Zugangswege und Zugriffsstufen bei Forschungsdatenzentren

Zugangswege und Zugriffstufen	Registrierung* und Autorisierung als wissenschaftliche/r Nutzer/in	Übermittlung zum/zur Datennutzer/in
Public Use File (Download, CD-Rom)	Nein	Ja (off-site-Nutzung)
Scientific Use File (Download, CD-Rom)	Ja	Ja (off-site-Nutzung)
Kontrollierte Datenfernverarbeitung (Remote Access)	Ja	Nein (off-site-Nutzung)
Gastwissenschaftlerarbeitsplatz	Ja	Nein (on-site-Nutzung)

* Die Registrierung beinhaltet die Zustimmung zu den speziellen Nutzungsbedingungen.

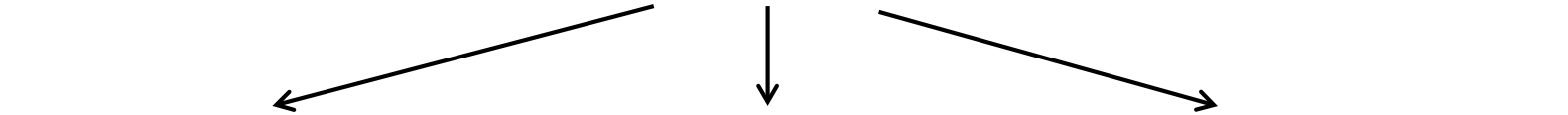
Meyermann, A. & Porzelt, M. (2019, S. 31). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

Hinweise

→ Nutzen Sie auch das vorher erstellte Skalenhandbuch

```
View(data)
str(data)
summary(data$variablename)
table(data$variablename)

# with R-package misty
freq(data$variablename) # frequency table
crosstab(data[, c("variable1", "variable2")]) # cross table
```



a) direkte Identifier **b) indirekte Identifier** **c) besondere Kategorien**

datenschutzrechtlich relevante Variablen

a) direkte Identifier
= personenbezogen

- Vor- und Nachnamen
- Adressen
- Geburtstage
- Versicherungsnummern
- Telefonnummern
- Email-Adressen
- …

b) indirekte Identifier
= personenbeziehbar

- Vornamen
- Stadt
- Straße
- Schulnamen
- Postleitzahlen
- Berufsangaben
- Bilder
- Datumsangaben
- Sprache zu Hause
- Geburtsort
- Seltene Merkmale (n≤5)

c) besondere Kategorien personenbezogener Daten

- Ethnizität
- Politische Einstellungen
- Religion
- Sexuelle Orientierung
- Medizinische und genetische Angaben

→ Erfassung dieser Merkmale muss in Einwilligungserklärung beschrieben werden

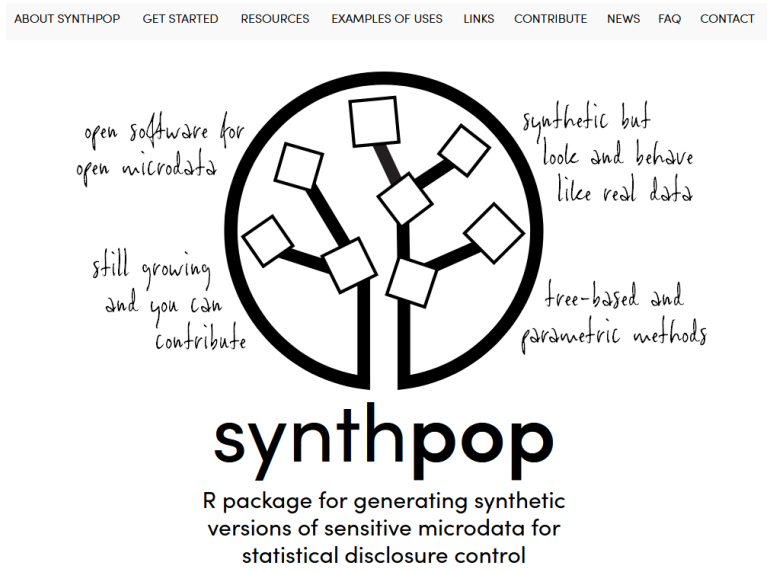
Meyermann, A. & Porzelt, M. (2019). Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung – eine Handreichung. *forschungsdaten bildung informiert 6*, Version 2.

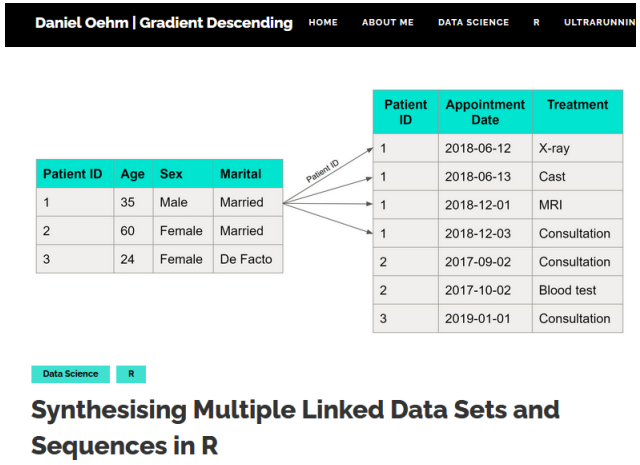
R-Paket synthpop *(Nowok, Raab & Dibben, 2016)*

- es werden neue Datensätze erzeugt, die Zusammenhänge zwischen Variablen in den Originaldaten annähernd abbilden (aber aus komplett künstlichen Fällen bestehen)
- Paket wird angewendet, um britische Studien und Zensus-Daten für Sekundäranalysen bereitzustellen, z. B. Scottish Longitudinal Study
- Ein Nachteil: Abbildung hierarchischer Datensätze nur begrenzt möglich

ERSTELLUNG SYNTHETISCHER DATEN

- eignet sich am besten ab einer Stichprobengröße von n = 500
- umfasst ebenfalls eine Codebook-Funktion: codebook.syn()
- umfangreiche Dokumentation: https://synthpop.org.uk/index.html
- Beispiel für hierarchische Daten: http://gradientdescending.com/synthesising-multiple-linked-data-sets-in-r/





ERSTELLUNG SYNTHETISCHER DATEN

Syntax: Datensatzerstellung (1)

```
#install.packages("synthpop") # install package for synthetic data
library(synthpop) # load package for synthetic data

# data preparation
####################

my.seed <- 1234 # fixing seed to make synthesized data reproducible

synthlist <-names(data) # put all variable names in one vector
exclude <- c("idstud", "idschool",  "idclass", "pptotwgt") #define vector with variables
that should be excluded from data synth (= not synthesized)

# define visiting-Sequence
visit.sequence.pisa <- setdiff(synthlist, exclude) # exclude variable names in "exclude"
-vector

# generate predictor matrix
sds.ini <- syn(data = data, visit.sequence = visit.sequence.pisa,
               drop.not.used = FALSE, m = 0, seed = my.seed)

#View(sds.ini$predictor.matrix)

# change predictor matrix
predictor.matrix.corrected <- sds.ini$predictor.matrix

# variable "pptotwgt" is will not be synthezised, but should be use as predictor for
synthesizing other variables
var_vorhersagen <- visit.sequence.pisa
```

ERSTELLUNG SYNTHETISCHER DATEN

Syntax: Datensatzerstellung (2)

```
# synthetisize data
###################

data.synthetic <- syn(data = data, visit.sequence = visit.sequence.pisa,
                      drop.not.used = FALSE, predictor.matrix = predictor.matrix.corrected,
                      m = 1, proper = FALSE, seed = my.seed)


pisa.data.synthetic <- data.synthetic$syn # extract and name synthetic data set
```

ERSTELLUNG SYNTHETISCHER DATEN

Syntax: Vergleich mit Originaldaten

```
# compare original and synthetic data

## univariate statistics

### all variables
compare(data.synthetic, data, vars = visit.sequence.pisa,
        nrow = 3, ncol = 5,  cols = c("black","ghostwhite"))

## regression: Dependent variable = mathematics achievement at t2

# original data
lm.original <- lm(ma_wle_t2 ~ ma_wle_t1 + grade_ma_t1 + schtype + gender + migration +
hisei + books + homepos, data = data)  # linear model, real data
summary(lm.original)
# Synthetic data
lm.synth <- lm.synds(ma_wle_t2 ~ ma_wle_t1 + grade_ma_t1 + schtype + gender + migration +
hisei + books + homepos, data = data.synthetic)  # linear model, synth data
summary(lm.synth)
lm.synthetic <- lm(ma_wle_t2 ~ ma_wle_t1 + grade_ma_t1 + schtype + gender + migration +
hisei + books + homepos, data = pisa.data.synthetic)  # linear model, synthetic data
summary(lm.synthetic)

# compare coefficients
theme_set(theme_bw())
compare(lm.synth, data, plot = "coeff", plot.intercept = TRUE, lcol = c("black","grey"))
```

FAZIT

- die Bereitstellung von Primärdaten nimmt stetig zu, ist manchmal sogar obligatorisch
- eine gute Datendokumentation ist der Schlüssel für erfolgreiche Sekundäranalysen, aber auch für die Re-Analyse und Reproduzierbarkeit Ihrer eigenen Daten/Ergebnisse
- Sie können Checklisten für die Datenaufbereitung in Ihrem Projekt einsetzten
- Sie können automatische Funktionen zur Erstellung von Skalenhandbüchern verwenden
- Sie können synthetische Daten als Anonymisierungsstrategie verwenden

- die Änderung der Arbeitsgewohnheiten und Arbeitsabläufe wird einige Zeit in Anspruch nehmen!

- Studie von Artner et al. (2020): Reproduzierbarkeit von Ergebnissen in der psychologischen Forschung: https://doi.org/10.1037/met0000365
  - geringe Reproduzierbarkeit von Forschungsergebnissen hängt neben Verfügbarkeit der Daten auch von deren Dokumentation und der Nachvollziehbarkeit der Auswertungssyntaxen ab
  - Reproduzierbarkeit von Forschungsergebnissen ist häufig aufwendig, insbesondere, wenn Primärforscher*innen nicht kontaktiert werden sollen
  - Empfehlung: kommentierte Auswertungssyntaxen zur Verfügung stellen

- Allgemeine Hinweise zur Dokumentation der Auswertungssyntaxen
  - einfache Ordnerstruktur mit Archiven und ggf. ReadMe-Dateien erstellen
  - Syntaxen so ausführlich wie möglich kommentieren und Gliederungspunkte integrieren (in RStudio: [strg] + [shift] + [R])
  - Versionen der statistischen Software in der Syntax mit angeben → in R über den Befehl sessionInfo()

- Reproduzierbarer Code mit Versionskontrolle in R
  - Versionskontrolle und Änderungsnachverfolgung mit git/github:
    - (1) R-Projekt erstellen
    - (2) usethis::use_git() und usethis::use_github()
  - Code Container (z. B. Code Ocean, Epskamp, 2019): Software Tools, die sicherstellen, dass alle Analysen immer mit derselben zugrundeliegenden Architektur durchlaufen – Analysen werden in einer Cloud durchgeführt, die gleiche Infrastruktur gewährleistet (Nachteil: ggf. datenschutzrechtlich bedenklich)
  - dynamische Berichte z. B. mit Rmarkdown: Der Ergebnisbericht beinhaltet auch die Syntaxen, mit der Ergebnisse erzeugt wurden

Hilfreiche Materialien

- Reproduzierbarer Workflow mit R (Brandmaier, 2020):

https://brandmaier.github.io/reproducible-data-analysis-materials/BioPsy2020.html#1

- dazugehöriges Symposium als Video: https://www.youtube.com/watch?v=nTVcMDVlyOI

- Tutorial von Mine Çetinkaya-Rundel (2020)
  - Präsentation: https://mine-cetinkaya-rundel.github.io/improve-repro-workflow-reproducibilitea-2020/slides/improve-repro-workflow-reproducibilitea-2020.pdf
  - Videotutorial: http://riotscience.co.uk/reproducible-data-science-with-r/

- allgemeine Hinweise zu „Clean Code“ (Martin et al. 2009):

https://enos.itcollege.ee/%7Ejpoial/oop/naited/Clean%20Code.pdf

WEITERE MATERIALIEN

Open Science und gute wissenschaftliche Praxis	Forschungsdatenmanagement	Recht & Datenschutz
Wie FAIR sind Deine Forschungsdaten? 10.5281/zenodo.2547339	Was sind Datenmanagementpläne? https://doi.org/10.18450/dataman/91	RatSWD-Handreichung Datenschutz https://www.konsortswd.de/aktuelles/publikation/handreichung-datenschutz-2/
Open Science Handbook https://open-science-training-handbook.gitbook.io/book/introduction	Informationsportale zum Thema Forschungsdatenmanagement (FDM) https://www.forschungsdaten.info/ https://www.forschungsdaten.org/ Watchlist mit FDM-Videos: https://av.tib.eu/watchlist/14442	Gutachten zu den rechtlichen Rahmenbedingungen des Forschungsdatenmanagements https://tu-dresden.de/gsw/jura/igetem/jfbimd13/ressourcen/dateien/dateien/DataJus/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf?lang=de
DFG-Kodex Gute wissenschaftliche Praxis https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf	Checkliste zur Erstellung eines Datenmanagementplans in der empirischen Bildungsforschung https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdbinfo_2.pdf	Datenschutzrechtliche Anforderungen in der empirischen Bildungsforschung - eine Handreichung https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-6_v2.pdf

WEITERE MATERIALIEN

Datzenschutz

- Meyer (2017) – Practical Tips for ethical data sharing
- Rocher et al. (2019) – Estimating the success of re-identifications in incomplete datasets using generative models

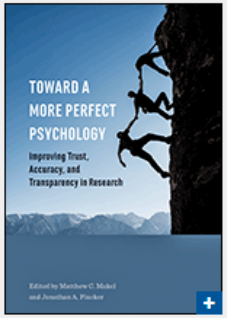
Synthetische Daten

- Nowok et al. (2016, 2017)
- Raab et al. (2017) – Guidelines for producing useful synthetic data
- Drechsler (2011) – Synthetic datasets for statistical disclosure control

WEITERE MATERIALIEN

Open Science Literatur

Toward a More Perfect Psychology:
Improving Trust, Accuracy, and Transparency in Research



Edited by Matthew C. Makel, PhD, and Jonathan A. Plucker, PhD

Pages: 323
Item #: 4318147
ISBN: 978-1-4338-2754-9
Copyright: 2017
Format: Softcover
Availability: In Stock
Also available on: Amazon Kindle

Examine or adopt this book for teaching a course

View sample pages from this book (PDF, 453KB)

Original Article | Published: 08 January 2021

Open Science als Beitrag zur Qualität in der Bildungsforschung

The contribution of Open Science to the quality of educational research

Georg Krammer & Erich Svecnik

Zeitschrift für Bildungsforschung **10**, 263–278(2020) | Cite this article

519 Accesses | Metrics

OPEN SCIENCE PUBLIKATIONEN

Nonlinear Relations Between Achievement and Academic Self-Concepts in Elementary and Secondary School: An Integrative Data Analysis Across 13 Countries

Draft 08/08/2019. This paper has been submitted for review but has not yet been peer reviewed. Please do not copy or cite without author's permission.

Lena Keller

Department of Educational Sciences, University of Potsdam, Potsdam, Germany,

Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany;

Franzis Preckel

Department of Psychology, University of Trier, Trier, Germany;

Martin Brunner

Department of Educational Sciences, University of Potsdam, Potsdam, Germany.

https://psyarxiv.com/8z563/

Check for updates

Research Article

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

It's Not Only Who You Are but Who You Are With: High School Composition and Individuals' Attainment Over the Life Course

Psychological Science
1–12
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797618794454
www.psychologicalscience.org/PS
SAGE

Richard Göllner[1], Rodica Ioana Damian[2], Benjamin Nagengast[1], Brent W. Roberts[1,3], and Ulrich Trautwein[1]
[1]Hector Research Institute of Education Sciences and Psychology, University of Tübingen; [2]Department of Psychology, University of Houston; and [3]Department of Psychology, University of Illinois at Urbana–Champaign

Open Practices

Mplus codes for univariate, multivariate, and mediation models, including output files, are publicly available on the Open Science Framework (https://osf.io/hksjv). The design and analysis plans for the experiments were not preregistered. The complete Open Practices Disclosure for this article can be found at http://

https://osf.io/nmtp6/

VIELEN DANK FÜR IHRE TEILNAHME AN UNSEREM WORKSHOP!