Bericht

Management und Bereitstellung von Forschungsdaten in der Psychologie: Überarbeitung der DGPs-Empfehlungen

DGPs-Kommission „Open Science“ (beschlossen durch den Vorstand der DGPs am 26.06.2020)

Mario Gollwitzer¹, Andrea Abele-Brehm², Christian J. Fiebach³, Roland Ramthun⁴, Anne Scheel⁵, Felix Schönbrodt¹ und Ulf Steinberg⁶

¹Ludwig-Maximilians-Universität München

²Friedrich-Alexander-Universität Erlangen-Nürnberg

³Goethe-Universität Frankfurt am Main

⁴Leibniz-Institut für Psychologie (ZPID), Trier

⁵Eindhoven University of Technology, Niederlande

⁶Technische Universität München

Zusammenfassung

Die Bereitstellung von Forschungsdaten im Rahmen von wissenschaftlichen Publikationen und öffentlich geförderten Forschungsprojekten wird heutzutage als ein zentraler Aspekt offener und transparenter Wissenschaftspraxis angesehen und von immer mehr Förderinstitutionen und wissenschaftlichen Fachzeitschriften eingefordert. Forschende sollten sich daher bemühen, die sogenannten FAIR-Prinzipien zu erfüllen, d.h. Forschungsdaten sollten auffindbar, zugänglich, interoperabel und wiederverwendbar sein. Ein systematisches Forschungsdatenmanagement unterstützt diese Ziele und soll es gleichzeitig ermöglichen, diese effizient zu verwirklichen. Mit den vorliegenden überarbeiteten Empfehlungen zum Management und zur Bereitstellung von Forschungsdaten spezifiziert die Deutsche Gesellschaft für Psychologie (DGPs) wichtige grundlegende Prinzipien des Datenmanagements in der Psychologie und greift hierbei auch Rückmeldungen der DGPs-Mitglieder auf, die im Rahmen einer Umfrage aus dem Jahr 2018 gewonnen wurden.

In einem ersten Schritt werden auf Basis disziplinspezifischer Definitionen von Roh-, Primär-, Sekundär- und Metadaten Empfehlungen hinsichtlich des Bearbeitungsgrades der im Rahmen einer Datenveröffentlichung bereitzustellenden Daten gegeben. In der Folge werden datenschutz- sowie urheber- und nutzungsrechtliche Aspek-

te der Datenbereitstellung diskutiert, bevor die qualitativen Anforderungen an vertrauenswürdige Repositorien zur Bereitstellung von Forschungsdaten definiert werden. Ausführlich wird in der Folge auf pragmatische Aspekte der Datenbereitstellung eingegangen, etwa auf die Unterschiede zwischen Datenveröffentlichungen vom Typ 1 und Typ 2, auf Nutzungssembargos, auf die Definition des „scientific use“ durch Nachnutzerinnen und -nutzer von veröffentlichten Daten sowie Empfehlungen zum Umgang mit etwaigen Konfliktfällen.

Besonders hervorzuheben ist die neue Empfehlung unterschiedlicher Zugriffsklassen für die Bereitstellung von Daten mit unterschiedlichen datenschutzrechtlichen oder forschungsethischen Anforderungen. Diese reichen von komplett offenen Daten ohne Nutzungsbeschränkungen („Zugriffsklasse 0“) über an standardisierte Bedingungen (z.B. die Nachnutzung für rein wissenschaftliche Zwecke) geknüpfte Datenbereitstellung („Zugriffsklasse 1“) über individualisierte Nutzungsverträge („Zugriffsklasse 2“) bis hin zu einem gesicherten Datenzugriff, der nur unter streng kontrollierten Bedingungen (z.B. in einem Forschungsdatenzentrum) erfolgen kann („Zugriffsklasse 3“). Die Umsetzung dieser wichtigen Neuerung ist jedoch geknüpft an technische Implementationen des Zugriffsklassenkonzeptes durch Datenrepositorien, die die entsprechenden Funktionalitäten bereitstellen.

Zusammenfassend zielen die überarbeiteten Empfehlungen darauf ab, Forscherinnen und Forschern pragmatische Wege für den offenen und transparenten Umgang mit psychologischen Forschungsdaten aufzuzeigen und dabei auch strukturelle Herausforderungen eines für alle Beteiligten gewinnbringenden „data sharings“ anzusprechen und entsprechende Maßnahmen vorzuschlagen.

Forschungsdatenmanagement ist angesichts zunehmend anspruchsvoller Datenschutz-Richtlinien einerseits und gestiegener Erwartungen an die Zugänglichkeit, Auffindbarkeit und Nachnutzbarkeit von Daten andererseits zu einer komplexen Aufgabe geworden – einer Aufgabe, deren Bedeutung für die Qualität und die Nützlichkeit von Wissenschaft allerdings kaum zu unterschätzen ist. Heute ist sich die wissenschaftliche Gemeinschaft weitgehend einig darin, dass Forschungsdaten grundsätzlich vier Kriterien erfüllen müssen (die FAIR-Kriterien¹; s. Wilkinson et al., 2016): Sie müssen auffindbar („Findable“), zugänglich („Accessible“), interoperabel („Interoperable“; d.h. mit anderen Daten integrierbar und von möglichst vielen Analyse- und Visualisierungsapplikationen nutzbar) sowie wiederverwendbar („Re-usable“) sein. Diese Prinzipien finden sich nicht nur in den Leitlinien der Deutschen Forschungsgemeinschaft (DFG) zur Sicherung guter wissenschaftlicher Praxis² sowie in den Berufsethischen Richtlinien der DGPs bzw. des BDP³ (s. dort Abschn. 7.3: „Grundsätze für Forschung und Publikation“, Absatz 14), sie werden auch von vielen anderen Förderinstitutionen (wie etwa dem European Research Council⁴) für die Mittelvergabe sowie zunehmend von wissenschaftlichen Fachzeitschriften im Zuge der Publikation von Fachartikeln vorausgesetzt. Gleichzeitig setzen die Europäische Datenschutz-Grundverordnung bzw. die aus ihr folgenden nationalen Datenschutzgesetze einer unrestringierten Bereitstellung und Nachnutzung von Forschungsdaten klare Grenzen.

Aber auch die Perspektive und die Interessen all jener Forscherinnen und Forscher, die ihre Daten zur Verfü-

gung stellen und die hierfür nötige Zeit und Arbeit investieren (wir werden im Folgenden von „Datenbereitstellenden“ sprechen⁵), muss berücksichtigt und gewürdigt werden. Datenbereitstellung und Datennachnutzung sind im Idealfall eine „Win-win-Situation“, von der die gesamte wissenschaftliche Gemeinschaft profitiert. So ist es notwendig, dass die wissenschaftliche Gemeinschaft die Bereitstellung von Forschungsdaten als eine wichtige wissenschaftliche Leistung wertschätzt und entsprechende Anreizsysteme entwickelt.

Angesichts dieser Komplexität hat der Vorstand der Deutschen Gesellschaft für Psychologie (DGPs) bereits im Jahre 2016 Empfehlungen zum Umgang mit Forschungsdaten in der Psychologie herausgegeben⁶. In der Präambel dieses Dokuments wurde festgelegt, dass die Empfehlungen „...nach fünf Jahren evaluiert und gegebenenfalls überarbeitet werden“ sollen (Schönbrodt, Gollwitzer & Abele-Brehm, 2017; S. 21). Hiermit wurde im Herbst 2018 die DGPs-Kommission „Open Science“ beauftragt⁷. Die nun vorliegende Überarbeitung der Datenmanagement-Empfehlungen bezieht sowohl informelle Rückmeldungen aus der DGPs-Mitgliederschaft als auch Befunde einer systematischen Evaluation dieser Empfehlungen (Abele-Brehm, Gollwitzer, Steinberg & Schönbrodt, 2019; Gollwitzer, Schönbrodt, Steinberg & Abele-Brehm, 2018) sowie aktuelle Stellungnahmen und Diskussionen in Deutschland (vgl. etwa RatSWD, 2018) und anderen Ländern (z. B. Sim et al., 2020) mit ein.

Die Kommission hielt es darüber hinaus für ratsam, ihre Überlegungen immer eng am technisch Machbaren auszurichten (und umgekehrt technische Lösungen an „best practices“ zu orientieren). Die aktuellen Empfehlungen sind daher in enger Abstimmung mit dem Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) an der Universität Trier entstanden. Das ZPID betreibt seit 2002 eine Plattform zur Archivierung und Bereitstellung von Forschungsdaten (ehemals „PsychData“, künftig „PsychArchives“⁸). Aufgrund der Kompatibilität mit den Vorgaben der Datenschutz-Grundver-

¹ https://www.go-fair.org/fair-principles_

² https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp.pdf

³ <https://www.dgps.de/index.php?id=85#c2001839>

⁴ https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf

⁵ Der Begriff „Datenbereitstellende“ wird hier für all jene verwendet, die ihre Primärdaten zur Nachnutzung zur Verfügung stellen, wobei uns sehr wohl bewusst ist, dass einer solchen „Bereitstellung“ im Regelfall ein immenser Aufwand (einschl. der Planung und Durchführung einer Studie, der Aufbereitung der Daten und der Zusammenstellung der Metadaten) vorausgeht, der durch den Begriff „Datenbereitstellende“ nicht wirklich abgebildet wird.

⁶ https://www.dgps.de/fileadmin/documents/Empfehlungen/Datenmanagement_deu_9.11.16.pdf (deutsche Version) bzw. https://www.dgps.de/fileadmin/documents/Empfehlungen/Data_Management_eng_9.11.16.pdf (englische Version)

⁷ Dieser Kommission gehören neben den Autorinnen bzw. Autoren der ursprünglichen Empfehlungen (Andrea Abele-Brehm, Mario Gollwitzer, Felix Schönbrodt) Christian Fiebach, Anne Scheel, Ulf Steinberg sowie – als Gäste – Mitarbeiter des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) und Vertreter der Arbeitsgruppe „Open Science“ der Psychologie Fachschaften Konferenz (PsyFaKo) an.

⁸ <https://leibniz-psychology.org/angebote/archivieren/>

ordnung der Europäischen Union einschließlich nationaler Datenschutzgesetze sowie einer klar disziplin-spezifischen Ausrichtung auf die Psychologie wurde die technische Umsetzbarkeit der hier formulierten Empfehlungen anhand von PsychArchives überprüft und sichergestellt.

1. Revision der Datenmanagement-Empfehlungen: Grundlegendes

Nach wie vor besteht das Ziel der DGPs-Empfehlungen zum Datenmanagement darin, praktische Lösungen zu entwickeln, die sich im Forschungsalltag als hinreichend effektiv, effizient, nachhaltig und akzeptabel erweisen und dabei den Interessen jener Personen, die Daten erheben und bereitstellen, gerecht werden, den aktuellen Datenschutzerfordernissen entsprechen sowie den gegenwärtigen Standards von Offenheit und Transparenz Rechnung tragen. Dabei sind Offenheit und Transparenz („Open Science“) unserer Ansicht nach kein Selbstzweck, sondern zum Zwecke des Qualitätsmanagements (d.h. der Vermeidung von Fehlern und Fehlschlüssen) sowie der Effizienzsteigerung wissenschaftlicher Prozesse essentiell. Die vorliegende Revision ergänzt, aktualisiert und präzisiert die Datenmanagement-Empfehlungen aus dem Jahr 2016 in Bezug auf

1. die Definition der Begriffe „Rohdaten“, „Primärdaten“ und „Sekundärdaten“ (s. Abschn. 2),
2. die mit der Bereitstellung von Daten verbundenen rechtlichen Erwägungen einschließlich Datenschutz, Urheber- und Nutzungsrecht (s. Abschn. 3),
3. Anforderungen an geeignete Repositorien (s. Abschn. 4),
4. den Zeitpunkt, zu dem eine Bereitstellung von Daten sinnvoll erscheint (s. Abschn. 5),
5. Zugriffs- und Nutzungsrechte Dritter (s. Abschn. 6) sowie
6. strukturelle Herausforderungen und Anreize sowie den Umgang mit Konfliktfällen (s. Abschn. 7).

Viele DGPs-Mitglieder, die im Jahr 2018 an unserer Umfrage zur Bekanntheit, Akzeptanz und Nützlichkeit unserer Empfehlungen teilgenommen haben (Abele-Brehm et al., 2019; Gollwitzer et al., 2018), haben sich insbeson-

dere Präzisierungen bezüglich der folgenden Themen gewünscht: Rechte von Datenbereitstellenden, Regelungen bezüglich Koautorenschaft und Konfliktfällen, Datenschutz und Urheberrecht sowie Möglichkeiten einer eingeschränkten Bereitstellung von Daten bzw. eine Definition spezieller Konditionen für die Nachnutzung bereitgestellter Daten. Diesen Wünschen tragen wir hier Rechnung.

2. Begriffsdefinitionen: Primärdaten, Metadaten, Sekundärdaten

2.1 Primärdaten

Da in der Folge häufig von „Primärdaten“ die Rede ist, soll dieser Datentyp zunächst definiert und von dem Begriff der Rohdaten abgegrenzt werden⁹. Zusätzlich wird häufig zwischen Primär- und Sekundärdaten unterschieden; diese Unterscheidung wird weiter unten (s. Abschn. 2.3) aufgegriffen. *Rohdaten* sind die Ursprungsaufzeichnungen, also etwa Kreuze auf einem Papierfragebogen, Zeichnungen, Audio- oder Videoaufnahmen, Blickbewegungsmessungen oder neuro- oder peripherphysiologische Aufzeichnungen (z. B. EEG, Herzrate). Rohdaten können daher definiert werden als die erste „nicht-flüchtige“ Form der Daten. Oft, aber nicht notwendigerweise, liegen Rohdaten bereits in digitalisierter Form vor. Als *Primärdaten* definieren wir für die psychologische Forschung die erste Übertragung der Rohdaten in ein digitales Format, also z. B. den Code „1“ für eine Ja-Antwort in einem Fragebogen. Häufig sind Rohdaten und Primärdaten äquivalent, etwa wenn die Antworten der Probandinnen und Probanden mittels einer Experimentalsteuerungssoftware oder im Rahmen eines Online-Survey erhoben und sofort digital gespeichert wurden. Somit sind Primärdaten in der Psychologie unbearbeitete (d.h. untransformierte, nicht-aggregierte etc.), in digitaler Form vorliegende quantitative oder qualitative Daten, z. B.

- bei Experimenten alle manipulierten und gemessenen Variablen für jeden Experimentaldurchgang jeder Person;
- bei Fragebögen die Antworten jeder Person auf jedem Item;

⁹ Diese Unterscheidung ist oft disziplinspezifisch. Allgemeiner unterscheidet beispielsweise die Nationale Forschungsdateninfrastruktur (NFDI) zwischen „Forschungsprimärdaten“ und „Forschungsdaten“, wobei erstere als „erhobene Rohdaten, die weder bearbeitet noch kommentiert bzw. mit Metadaten versehen wurden“ definiert sind. Forschungsdaten sind in diesem Sinne alle Daten, Sekundäranalysen, Visualisierungen, Ergebnisse, etc., die im Forschungsprozess entstehen (<https://www.forschungsdaten.info/praxis-kompakt/glossar/#c269824>). Die DFG definiert Forschungsprimärdaten als „Daten, die im Verlauf von Quellenforschungen, Experimenten, Messungen, Erhebungen oder Umfragen entstanden sind. Sie stellen die Grundlagen für die wissenschaftlichen Publikationen dar“ (https://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf). Auch hier wird allerdings auf disziplinspezifische Definitionen hingewiesen.

- bei Freitext-Eingaben der Originalwortlaut;
- digitalisierte Videoaufnahmen;
- Downloads oder Screenshots von Inhalten sozialer Medien (z.B. Facebook-Profil oder Twitter-Nachrichten);
- bei (neuro)physiologischen Daten (wie EEG- oder fMRT-Daten) verlustfrei umgewandelte Daten in einem standardisierten Rohdatenformat (z.B. EDF, DICOM oder NIFTI), die nicht aggregiert sind und nicht auf nur wenige „regions of interest“, Messkanäle oder ähnliches beschränkt sind¹⁰.

Primärdaten beinhalten auch die Daten jener Personen, die von der Datenanalyse ausgeschlossen wurden (außer wenn der Ausschluss darauf beruht, dass die teilnehmende Person während oder nach der Datenerhebung ihre Einwilligung zurückgezogen hat). Zusammengefasst definieren wir Primärdaten als die Menge aller Datenpunkte, die im Rahmen einer Studie bzw. eines Projekts erhoben wurden, in ihrer ersten digitalen Übertragung, aber ansonsten in völlig unbearbeiteter Form.

Primärdaten sollten in Form von offenen bzw. frei zugänglichen Datenformaten veröffentlicht werden (um „accessibility“ und „interoperability“ im Sinne der FAIR-Prinzipien zu garantieren). Wenn möglich, sollten Daten in einer standardisierten Dateistruktur zur Verfügung gestellt werden, da auch dies die Möglichkeiten der Nachnutzung erleichtert. Die Brain Imaging Data Structure¹¹ als standardisierte Dateiablagestruktur für MRT-Daten kann hier als „best practice-Beispiel“ dienen.

2.2 Metadaten

Primärdaten erfordern *eindeutige* und *reichhaltige Metadaten*, um ihre Auffindbarkeit und ihre Nachnutzbarkeit zu gewährleisten. Metadaten stellen strukturierte Informationen über den Primärdatensatz bereit; sie beinhalten all jene Informationen, die nötig sind, um die Ergebnisse und Interpretationen der Studie, für die die Primärdaten erhoben wurden, zu überprüfen und nachzuvollziehen und um einschätzen zu können, für welche Art der Nachnutzung sich ein Primärdatensatz eignet. Hierzu gehö-

ren nicht nur ein „data dictionary“ (d.h. eindeutige Beschriftungen und Erklärungen der Variablen im Datensatz), sondern auch Informationen über die untersuchte Stichprobe (z.B. Art der Rekrutierung, Feldzugang, Ziehung der Stichprobe), den Prozess der Datenerhebung, -verarbeitung und -analyse sowie darüber, wer die Daten wie generiert hat (RatSWD, 2018).

Metadaten können inhaltlicher oder technisch-administrativer Art sein. Zu den technischen Metadaten gehören Informationen über alle relevanten Parameter der Datenerhebung (bei EEG-Daten also bspw. die Sampling-Rate sowie die Filterung). Diese Informationen sind notwendig, um sowohl die Vergleichbarkeit etwaiger empirischer Replikationen als auch die Möglichkeit der Datenanalyse im Rahmen der Nachnutzung (FAIR-Prinzip „reusability“) sicherzustellen. Diese Informationen können in Form von Verweisen in den Rohdatensätzen oder als separate Begleitdokumente (d.h. als Textdatei) im Repository zur Verfügung gestellt werden. Mit dem Ziel der Interoperabilität wurden übergreifende Metadaten-Standards etabliert, etwa die *Data Documentation Initiative*¹² oder die *Dublin Core Metadata Initiative*¹³. Disziplinspezifische Informationen finden sich beispielsweise auf der Webseite des Datenmanagementsystems DataWiz, welches vom ZPID speziell für die psychologische Forschung entwickelt wurde.¹⁴

2.3 Sekundärdaten

Als *Sekundärdaten* werden Daten verstanden, die bereits ersten Verarbeitungsschritten unterzogen wurden¹⁵. Hierzu zählen beispielsweise Transformationen von Variablen, Aggregation einzelner Beobachtungen, Berechnung von Skalenwerten, Berechnung von ereigniskorrelierten Potentialen durch Mittelung, oder der Ausschluss von Werten, von denen angenommen wird, dass sie die Robustheit der Analyseergebnisse beeinträchtigen (z.B. Extrem- und Ausreißerwerte, ungültige Fälle). Im Sinne der Reproduzierbarkeit und Nachnutzbarkeit von Daten sollten immer Primärdaten bereitgestellt werden; allerdings kann es aus rechtlichen oder forschungsethischen Gründen notwendig sein, anstelle der Primärdaten die aus ih-

¹⁰ Für MRT-Daten s. z.B. die „Best Practices in Data Analysis and Sharing in Neuroimaging using MRI“ der Organization for Human Brain Mapping (<http://www.humanbrainmapping.org/files/2016/COBIDASreport.pdf>). Für EEG- und MEG-Daten s. z.B. die „Best Practices in Data Analysis & Sharing in Neuroimaging using MEG“ (<https://cobidasmeeg.wordpress.com>).

¹¹ <https://bids.neuroimaging.io>; zusätzlich möchten wir auf das momentan im Aufbau befindliche Projekt „psych-DS“ verweisen, dessen Ziel es ist, eine an BIDS angelehnte Dateistruktur für psychologische Daten im Allgemeinen zu entwickeln, s. <https://github.com/psych-ds/psych-DS>

¹² <https://ddialliance.org>

¹³ <https://dublincore.org>

¹⁴ <https://lifp.de/psychologische-metadaten>

¹⁵ Andere Autoren definieren „Sekundärdaten“ anders, etwa als jene Daten, die zu einem anderen Zweck nachgenutzt werden als den, zu dem sie ursprünglich erhoben worden sind (vgl. Hox & Boeije, 2005).

nen hervorgegangenen Sekundärdaten bereitzustellen (s. Abschn. 3.2 und 3.3). Wenn möglich, sollte in diesem Falle der Programmcode zur Verfügung gestellt werden, auf dessen Basis nachvollzogen werden kann, wie der Sekundärdatensatz aus den Primärdaten generiert wurde.

3. Rechtliche Aspekte der Datenbereitstellung

3.1. Datenschutz

Datenschutzrechtliche Einschränkungen müssen bereits in der Planungsphase einer Untersuchung berücksichtigt werden und sind insofern auch für die spätere Datenbereitstellung essentiell. Hierbei sind die Datenschutzgrundverordnung (DSGVO)¹⁶ sowie für Hochschulen in der Regel die einschlägigen Paragraphen der Landesdatenschutzgesetze und für andere wissenschaftliche Einrichtungen diejenigen des Bundesdatenschutzgesetzes (BDSG) zu beachten. In den Regelungen in Deutschland finden sich jeweils datenschutzrechtliche Bestimmungen für die wissenschaftliche Forschung (z. B. § 27 BDSG oder § 17 DSG NRW), mit denen die Öffnungsklauseln der grundlegenden DSGVO genutzt werden (Art. 89). Weiterhin sind von besonderer Relevanz Art. 4 DSGVO (Begriffsbestimmungen), Art. 5 DSGVO (Grundsätze für die Verarbeitung personenbezogener Daten), Art. 7 DSGVO (Bedingungen für Einwilligungen) sowie Art. 32 DSGVO (Sicherheit der Datenverarbeitung). Weitere Hinweise finden sich in einschlägigen Publikationen (etwa RatSWD, 2020).

3.2 Personenbezogene Daten

Personenbezogene Daten (d. h. „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“; Art 4 Zif. 1 DSGVO) müssen frühest-

möglich anonymisiert bzw. hilfsweise pseudonymisiert¹⁷ werden. Damit muss sichergestellt werden, dass Personen nicht aus der Kombination verschiedener erhobener Merkmale – auch solcher, die in unterschiedlichen Studien mit den gleichen Teilnehmerinnen und Teilnehmern (z. B. Studierende im 1. Fachsemester Psychologie an der Universität XY) erhoben wurden – identifizierbar sind (Art. 89 DSGVO).

Datenschutzrechtliche Erwägungen spielen jedoch nicht nur auf der Ebene einzelner Personen, sondern auch auf relevanten Aggregatebenen eine Rolle: So muss speziell im Falle sensitiver Fragestellungen (z. B. illegales Verhalten, Suizidraten) besonders darauf geachtet werden, inwiefern einzelne Schulen, Firmen, etc. in den Daten bzw. durch eine Zusammenführung von Datensätzen eindeutig identifiziert werden können (RatSWD, 2020).

Im Regelfall ist vor Untersuchungsbeginn die Erstellung eines Verzeichnisses von Verarbeitungstätigkeiten erforderlich, in dem dokumentiert wird, welche personenbezogenen Daten auf welcher Grundlage und auf welche Art verarbeitet werden und welche technischen und organisatorischen Maßnahmen getroffen werden, um Datenschutz und Datensicherheit zu gewährleisten. Ist durch die Datenerhebung oder -verarbeitung ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zu erwarten, ist zusätzlich eine Datenschutz-Folgeabschätzung vorzunehmen (z. B. Martin, Friedewald, Schiering, Mester & Hallinan, 2020; s. a. die *Guidelines on Data Protection Impact Assessment* der EU¹⁸). Dies gilt insbesondere, wenn Daten gesetzlich besonders geschützter Art erhoben werden (z. B. ethnische Herkunft, politische, religiöse oder philosophische Überzeugungen, Angaben zum Sexualleben sowie bspw. physiologische Daten oder GPS-Daten aus Tracking-Geräten, die Rückschlüsse auf Gesundheitszustand oder sehr einfach eine Reidentifizierung zulassen).

Datenbereitstellende müssen dokumentieren, ob und wie die Daten anonymisiert/pseudonymisiert worden sind. Das Verzeichnis von Verarbeitungstätigkeiten muss auf Anfrage den zuständigen Aufsichtsbehörden bereitgestellt werden; im Fall hoch risikobehafteter Daten (s. o.) ist auch die Folgeabschätzung zu dokumentieren. Wenn

¹⁶ Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32016R0679>

¹⁷ Anonymisierte Informationen sind „Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.“ (Erwägungsgrund 26 DSGVO, teilweise übernommen in Landesdatenschutzgesetze). Pseudonymisierung meint „die Verarbeitung personenbezogener Daten in einer Weise, in der die Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die Daten keiner betroffenen Person zugewiesen werden können“ (Art. 4 Zif. 5 DSGVO).

¹⁸ https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236

all diese Informationen durch die Datenbereitstellen vorliegen, sollten diese rechtlich nicht mehr belangt werden können, falls Datennachnutzende im Zuge der Nachnutzung gegen Datenschutzrechte verstoßen.

3.3 Einwilligung

Nicht nur aus juristischen (RatSWD, 2020), sondern auch aus forschungsethischen Erwägungen heraus (Deutsche Gesellschaft für Psychologie DGPs, 2018) ist es entscheidend, dass Teilnehmende über den Nutzen, die Risiken und die Form der Datenerhebung sowie über die Zwecke der Datennutzung, -speicherung und -weiterverwendung in verständlicher Form aufgeklärt werden, so dass sie informiert und aufgrund freier Entscheidung in die Verarbeitung ihrer Daten einwilligen können. Die Freiwilligkeit der Teilnahme muss stets gewährleistet bleiben. Der Abbruch der Studienteilnahme durch Probandinnen und Probanden muss zu jeder Zeit möglich sein. Den Probandinnen und Probanden sollte die Möglichkeit zugesichert werden, Einblick in die erhobenen Daten zu nehmen und diese korrigieren zu dürfen (sofern die Daten nicht vollständig anonymisiert wurden). Bei Video- und Tonaufnahmen ist die Zustimmung zwingend erforderlich und muss entsprechend eingeholt werden; die Einholung der Einwilligung muss dokumentiert werden (vgl. Deutsche Gesellschaft für Psychologie DGPs, 2018)¹⁹.

Der Deutsche Ethikrat unterscheidet verschiedene Einwilligungsmodelle (s. RatSWD, 2020; S. 27). Bei der *Blanko-Einwilligung* stimmen Untersuchungsteilnehmende einer inhaltlich unbestimmten zukünftigen Nutzung und der Weitergabe der Daten zu. Bei der *Dynamischen Einwilligung* werden Untersuchungsteilnehmende wiederholt kontaktiert, um ihre informierte Einwilligung zu einzelnen Fragestellungen einzuholen. Sie stehen in regelmäßigem Kontakt mit einer Datenbank (z. B. über eine Online-Plattform oder telefonische Kontakte). Die *Kaskaden- oder Meta-Einwilligung* ist eine Erweiterung der dynamischen Einwilligung. Hierbei muss die Einwilligung nicht zwangsläufig für jede Fragestellung neu eingeholt werden, sondern die Untersuchungsteilnehmenden können zwischen verschiedenen Optionen wählen.

Schließlich wurde speziell für die wissenschaftliche Forschung in Erwägungsgrund 33 der DSGVO die Möglichkeit einer „breiten“ Einwilligung geschaffen. Der europäische Verordnungsgeber geht dabei davon aus, dass

es häufiger vorkommen kann, dass der Zweck der Verarbeitung personenbezogener Daten für Zwecke der wissenschaftlichen Forschung zum Zeitpunkt der Erhebung der personenbezogenen Daten nicht vollständig angegeben werden kann. Daher lässt er die Möglichkeit zu, dass Untersuchungsteilnehmende (a) für bestimmte Bereiche wissenschaftlicher Forschung oder für Teile von Forschungsprojekten, (b) in dem vom verfolgten Zweck zugelassenen Maß und (c) unter Einhaltung anerkannter ethischer Standards der wissenschaftlichen Forschung eine „breite“ Einwilligung erklären. Der „bestimmte Bereich“ muss dabei einen Zusammenhang mit dem ursprünglichen Forschungsziel haben.

Neben der grundsätzlichen Einwilligung müssen Teilnehmerinnen und Teilnehmer einer Studie darauf aufmerksam gemacht werden, dass ihre anonymisierten Daten ggf. für eine Nachnutzung durch Dritte zur Verfügung gestellt werden und dass Zweck, Art und Umfang dieser Nachnutzung zum gegenwärtigen Zeitpunkt noch nicht absehbar sind. Eine explizite Einwilligung zur Datennachnutzung ist in jedem Fall einzuholen, wenn diese Daten nicht vollständig anonymisiert werden können (Metschke & Wellbrock, 2002). Wenn Daten vollständig anonymisiert sind, ist die Einholung einer entsprechenden Einwilligung zwar rechtlich nicht erforderlich (da keine individuellen Zuordnungen mehr möglich sind), allerdings dennoch aus forschungsethischer Sicht geboten (Deutsche Gesellschaft für Psychologie DGPs, 2018).²⁰

Verweigern einzelne Teilnehmerinnen oder Teilnehmer ihre Zustimmung zu einer eventuellen Nachnutzung ihrer nicht vollständig anonymisierten Daten, dürfen die Daten dieser Teilnehmerinnen und Teilnehmer nicht veröffentlicht werden. Wenn Daten nicht veröffentlicht werden können, sollten die entsprechenden Gründe an geeigneter Stelle (z. B. in einer Fußnote in der Publikation und in der Dokumentation des Datensatzes im Repositorium) benannt werden. Solche Bedenken sollten jedoch umgekehrt nicht als Rechtfertigung dazu dienen, Forschungsdaten nicht zu veröffentlichen, obwohl es rechtlich und ethisch unproblematisch wäre. Darüber hinaus soll bei bestehenden rechtlichen Einschränkungen dargelegt werden, welche Arten von aggregierten Daten oder anonymisierbaren bzw. pseudonymisierbaren Teildatensätzen dennoch veröffentlicht werden können.

Um Versuchspersonen in der Praxis das Recht zu gewähren, ihre Daten einzusehen oder auch nach der Datenerhebung rückwirkend ihre Einwilligung zurückzu-

¹⁹ Vorlagen für Einwilligungserklärungen stellt etwa die Ethikkommission der DGPs auf der Internetseite des TransMIT-Zentrums für wissenschaftlich-psychologische Dienstleistungen (ZwpD) bereit: <https://zwpd.transmit.de/zwpd-dienstleistungen/zwpd-ethikkommission/vorlagen-antragstellung>

²⁰ Dies betrifft unter Umständen auch Daten, die im Rahmen früherer Studien erhoben wurden.

ziehen, kann ggf. für einen kurzen Zeitraum eine Liste geführt werden, die die Identifizierung der Versuchspersonen anhand vergebener Pseudonymisierungsschlüssel erlaubt. Nach Ablauf einer vorher festgelegten und den Versuchspersonen kommunizierten Frist (bspw. X Stunden/Tage nach Ende der Datenaufzeichnung) wird die Entblindungsliste, in der Pseudonymisierungsschlüsseln Klarnamen zugeordnet sind, gelöscht.

3.4 Urheberrecht und Nutzungsrecht

In der Regel sind Primärdaten nicht urheberrechtlich geschützt (Guibault & Wiebe, 2013; Hillegeist, 2012; Spindler & Hillegeist, 2011), allerdings müssen immer Fragen zu Nutzungsrechten geklärt werden, bevor eine Veröffentlichung dieser Daten erfolgt. Im Falle von Daten, die von wissenschaftlichen Beschäftigten einer Forschungseinrichtung erhoben wurden, liegen die entsprechenden Nutzungsrechte üblicherweise bei dieser Einrichtung²¹. Leider ist die Rechtslage im Bereich des Verhältnisses zwischen Arbeitgeber und Arbeitnehmer im Bereich Forschungsdaten sehr komplex, so dass in diesem Zusammenhang generelle Aussagen schwer möglich sind²². Insofern müssen Fragen zu Nutzungsrechten immer geklärt werden, bevor eine Veröffentlichung von Forschungsdaten erfolgt.

4. Anforderungen an ein geeignetes Repositorium

Die Primärdaten sollen in digitaler Form auf einem vertrauenswürdigen Repositorium bereitgestellt werden. Wichtige Qualitätsmerkmale eines vertrauenswürdigen Repositoriums sind:

- die wirtschaftliche Unabhängigkeit sowie die wissenschaftliche Professionalität der bereitstellenden Institution,
- die Zugänglichkeit der Daten (es muss möglich sein, die Daten kostenfrei und nach einem festen, abgestuften Zugriffsklassenmodell zu erhalten; s. Abschn. 6.2),
- die Zitierfähigkeit der Daten (es muss ein „Digital Object Identifier“ (DOI) vergeben werden, der sich auf eine eindeutige Version der Daten bezieht),
- die Wahrung des Datenschutzes entsprechend EU-Recht in dem Land, in dem das Repositorium die Daten speichert, die Klärung der Rechte an den Daten (mit dem Speichern der Daten darf kein Abtreten der ausschließlichen Nutzungsrechte an Dritte verbunden sein²³) und
- die Persistenz der Daten (d.h. es muss geklärt sein, wie lange und unter welchen Bedingungen die Daten im Falle der Auflösung des Repositoriums oder seiner betreibenden Institution weiter unbeschränkt zur Verfügung stehen).

Daher ist ein vertrauenswürdigen öffentliches Repositorium (z.B. PsychArchives des ZPID²⁴, datorium bei GESIS²⁵ oder ein gut ausgebautes universitäres Repositorium) einem Zeitschriftenrepositorium vorzuziehen. Von einer Bereitstellung auf privaten oder persönlichen universitären Webseiten wird dringend abgeraten. Zudem sollte die Institution, die das Repositorium bereitstellt, qualifizierte Information und ggf. Beratung bei der Speicherung der Primärdaten zur Verfügung stellen können. Bei der Wahl des Repositoriums sind Einschränkungen, die sich aus rechtlichen oder forschungsethischen Richtlinien ergeben (z.B. ein Verbot, Daten auf einem ausländischen oder außereuropäischen Server zu speichern), zu beachten.

²¹ In bestimmten Fallkonstellationen kann dies anders sein und die Nutzungsrechte liegen bei den Forschenden selbst; vgl. dazu Hillegeist (2012) und <https://blogs.hrz.tu-freiberg.de/oersax/urheberrecht-des-wissenschaftlichen-personals-an-hochschulen>.

²² In Bezug auf Ansprüche an Forschungsdaten können beispielsweise Urheberrecht, Leistungsschutzrecht, Patentrecht, Privatrecht (Arbeitsrecht) und Verfassungsrecht (Wissenschaftsfreiheit) eine Rolle spielen. Eine Übersicht bieten Hillegeist (2012), https://www.forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht_und und <https://www.forschung-und-lehre.de/wem-gehoren-forschungsdaten-1013>. Es scheint daher empfehlenswert, mit der Hochschulleitung Kontakt aufzunehmen und ggf. auf eine generelle Regelung im Umgang mit Forschungsdaten hinzuwirken, die sich zu Open Science-Prinzipien bekennt und die Veröffentlichung von Daten, die von Mitarbeitenden der Einrichtung erhoben und zusammengestellt wurden, durch diese erlaubt und fördert. Durch eine solche Forschungsdaten-Richtlinie entfallen Zweifelsfälle und Anfragen Forschender im Einzelfall. Beispiele für solche Forschungsdatenrichtlinien finden sich bei der TUM (<https://www.it.tum.de/projekte/forschungsdaten-management/>) und HHU (https://www.uni-duesseldorf.de/redaktion/fileadmin/redaktion/Fakultaeten/Medizinische_Fakultaet/MedRSD/Dok_GWP/Forschungsdaten_Richtlinie_2015.pdf).

²³ Das einfache Nutzungsrecht, d.h. das Recht zur Archivierung und Vervielfältigung, muss dem Betreiber des Repositoriums übertragen werden, damit der Betrieb des Repositoriums überhaupt möglich ist.

²⁴ <https://psycharchives.org>

²⁵ <https://data.gesis.org/sharing/#!Home>

5. Zeitpunkt und Form der Datenbereitstellung

In Bezug auf den Zeitpunkt der Datenveröffentlichung unterscheiden wir – wie auch schon in den ursprünglichen Empfehlungen (Schönbrodt et al., 2017) – zwei Typen der Datenbereitstellung.

5.1 Bereitstellung von Daten als Bestandteil einer Publikation („Typ 1“)

Mit Erscheinen einer Publikation soll die Person oder Gruppe, die die Daten erhoben hat (die Datenbereitstellenden), alle Primärdaten sowie die dazugehörigen Metadaten bereitstellen, die zur Reproduktion der publizierten Ergebnisse notwendig sind, unabhängig davon, in welchem Kontext (bspw. drittmittelfinanziertes Forschungsprojekt oder studentische Abschlussarbeit) diese entstanden sind. In den Begriff „Publikation“ sind explizit auch solche Manuskripte eingeschlossen, die der Öffentlichkeit bereits vor dem Erscheinen in einer Fachzeitschrift als sogenannte „Preprints“ zugänglich gemacht werden. Bei auf dem Standardweg publizierten Manuskripten – also solchen, die erst mit dem Erscheinen in einer Fachzeitschrift öffentlich werden – setzen viele Fachzeitschriften schon jetzt voraus, Primär- und Metadaten bereits während des Reviewprozesses für die Gutachter zugänglich zu machen. Verantwortlich für die langfristige technische und inhaltliche Nachvollziehbarkeit des Datensatzes ist – sofern nicht anders geregelt – die Erstautorin oder der Erstautor bzw. die korrespondierende Autorin oder der korrespondierende Autor.

Bei Simulationsstudien sollte der datengenerierende Code bereitgestellt werden. Sofern die simulierten Daten durch diesen Code nicht in hinreichend eindeutigem Maße wiederhergestellt werden können oder eine Wiederherstellung mit unverhältnismäßigem Aufwand oder Kosten verbunden wäre, kann es notwendig sein, auch die simulierten Daten selbst bereitzustellen. Die Dokumentation des Programmcodes sollte alle notwendigen Informationen zur Reproduktion der Daten umfassen.

In der Dokumentation zu den entsprechenden Primärdaten sollten im Regelfall auch solche Variablen genannt werden, die im Rahmen der jeweiligen Studie ebenfalls erhoben, in der betreffenden Publikation aber nicht berücksichtigt wurden (vgl. „standard reviewer disclosure request“²⁶). Die Daten zu diesen Variablen werden im

Primärdatensatz ergänzt, wenn sie entweder später selbst Teil einer Publikation sind oder wenn das Forschungsprojekt abgeschlossen ist und der Gesamtdatensatz bereitgestellt wird (Datenveröffentlichung Typ 2; s. Abschn. 5.2).

Vor der Veröffentlichung – idealerweise bereits vor der Datenerhebung – muss geklärt werden, ob alle Daten bereitgestellt werden können oder nur ein Teil von ihnen, und ob – etwa aus forschungsethischen Erwägungen – nur Sekundärdaten bereitgestellt werden können. Diese Erwägungen müssen dokumentiert werden (s. Abschn. 3.2 und 3.3).

5.2 Bereitstellung von Daten nach Projektabschluss („Typ 2“)

Gemäß den Empfehlungen der DFG soll der Gesamtdatensatz eines Projekts „unmittelbar nach Abschluss der Forschungen oder nach wenigen Monaten der Öffentlichkeit frei zur Verfügung gestellt werden“²⁷. Das beinhaltet auch alle relevanten Daten des Projekts, die zu diesem Zeitpunkt noch nicht Bestandteil einer Publikation sind, einschließlich der entsprechenden Metadaten (s. Abschn. 2.2).

Welche Daten als „relevant“ anzusehen sind, liegt im Ermessen der Projektverantwortlichen. Irrelevant sind beispielsweise Daten, die auf fehlerhaft programmierten Experimenten beruhen oder im Rahmen von hochgradig explorativen Pilotstudien erhoben wurden. Um dem Problem des Publikationsbias entgegenzuwirken, sollen in jedem Fall auch solche Daten bereitgestellt werden, die nicht-erwartungskonforme Ergebnisse erbracht haben (sofern nicht ein technischer oder handwerklicher Fehler hierfür verantwortlich gemacht werden kann). Wir halten es in diesem Zusammenhang für folgerichtig, Projektverantwortliche zu verpflichten, im Abschlussbericht eines Projekts explizit anzugeben, welche Studien durchgeführt wurden und wo sich die entsprechenden Primärdaten befinden (und wie sie ggf. abgerufen und genutzt werden können). Gutachterinnen und Gutachter sollten Abschlussberichte auch daraufhin bewerten, ob die Dokumentation der Datenbereitstellung hinreichend transparent und korrekt ist.

Der Zeitpunkt, zu dem ein Projekt als abgeschlossen gilt, ist von der Komplexität des Projekts abhängig. Bei drittmittelgeförderten Projekten gilt jedoch im Regelfall, dass die Einreichung des Abschlussberichts den Projektabschluss darstellt. Viele Förderinstitutionen verlangen,

²⁶ <https://osf.io/had3/>

²⁷ http://dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf

dass die Bereitstellung der Daten, die im Rahmen des Projekts erhoben wurden, so zeitnah wie möglich nach Abschluss des Projekts erfolgen soll. Vor diesem Hintergrund sollte allen Projektverantwortlichen klar sein, dass Forschungsdatenmanagement eine Aufgabe ist, die von Anfang an begleitend zum Projekt mit bedacht werden muss (und nicht erst gegen Abschluss des Projekts in Angriff genommen wird). Dennoch mag es Fälle geben, in denen eine Datenaufbereitung und -bereitstellung erst gegen Ende eines Projekts möglich ist und sich die Bereitstellung der Daten entsprechend verzögert. In diesem Fall ist es sinnvoll, entsprechende Datenstrukturen einschließlich der dazugehörigen Metadaten bereits im Repositorium anzulegen und eine Textdatei beizufügen, aus der der erwartete Zeitpunkt der Bereitstellung hervorgeht (s.a. Abschn. 5.3).

Die Veröffentlichung Typ 2 bezieht sich insbesondere auf Projekte, die einen definierten Erhebungsumfang und einen definierten Projektabschluss haben (so wie dies typischerweise bei einem drittmittelfinanzierten Forschungsprojekt der Fall ist). Bei kontinuierlich laufenden Projekten (z.B. aus universitären Mitteln) ist der genaue Zeitpunkt des Projektabschlusses möglicherweise schwer zu definieren. Allerdings gilt auch hier, dass die Datenbereitstellung ohne unnötige Verzögerungen erfolgen sollte.

5.3 Nutzungsembargo

Datenbereitstellende können ein Embargo für die Datennachnutzung definieren. Das bedeutet, dass die bisher noch nicht für Publikationen verwendeten Daten, die im Zuge eines Forschungsprojekts erhoben wurden, zwar – wie in Abschnitt 5.2 ausgeführt – zeitnah nach Beendigung des Projekts auf einem Repositorium abgelegt, aber erst zu einem späteren Zeitpunkt für Dritte zugänglich gemacht werden. Ein Nutzungsembargo kann beispielsweise gerechtfertigt sein, um die vorübergehende Sicherung des geistigen Eigentums zu ermöglichen und/oder die Datenbereitstellenden vor Nachteilen zu schützen. Ein Nutzungsembargo für die Datenveröffentlichung vom Typ 2 sollte in der Regel nicht länger als fünf Jahre nach Projektabschluss gelten. Embargos müssen generell begründet werden.

Daten, die als Teil einer Publikation veröffentlicht wurden (Datenveröffentlichung Typ 1), sollten in der Regel nicht mit einem Embargo belegt werden. In Ausnahmefällen (z.B. bei einer extrem aufwändigen Datenerhebung) können auch diese Daten mit einem Embargo belegt werden; in diesem Fall sind zusätzlich etwaige Regelungen der Fachzeitschrift, in der der entsprechende Artikel publiziert wurde, zu berücksichtigen. Das Embargo

im Zusammenhang mit einer Datenveröffentlichung vom Typ 1 sollte jedoch deutlich kürzer sein als die für die Datenveröffentlichung vom Typ 2 empfohlene Embargoperiode von maximal fünf Jahren. Im Falle eines Embargos bei einer Datenveröffentlichung vom Typ 1 muss sichergestellt sein, dass die Daten ab dem Zeitpunkt der Veröffentlichung der Publikation zumindest auf Anfrage für eine Reproduktion der berichteten Ergebnisse zur Verfügung stehen (s. Beispiel 3 in Abschn. 6.2).

Um ein Nutzungsembargo zu dokumentieren, eignet sich eine einfache Textdatei, die im Repositorium abgelegt ist, und die Informationen über die Art des Embargos, die Begründung, die Länge des Embargos sowie weitere Informationen zu den noch nicht bereitgestellten Daten (Metadaten; ggf. Verweis auf ein Codebuch oder ein „data dictionary“; s. Abschn. 2.2) enthält. In geeigneten Repositorien kann ein Nutzungsembargo auch realisiert werden, indem zunächst (und temporär) eine restriktivere Zugriffsklasse vergeben wird (s. Abschn. 6.2), wobei der Datensatz nach dem festgelegten Zeitraum automatisch auf eine weniger strenge Zugriffsklasse wechselt. Das bedeutet, dass die Daten nach Ablauf des Embargos zu den vorher festgelegten Bedingungen für die Nachnutzung verfügbar gemacht werden. Als „best practice“ für die Umsetzung eines Embargos empfehlen wir, ein „file only embargo“ zu nutzen, in dessen Rahmen die eigentlichen Datenfiles zwar mit einem Embargo belegt sind, die entsprechenden Metadaten aber bereits früher öffentlich zugänglich (und somit auffindbar) sind. Außerdem sollte die Aufhebung des Nutzungsembargos nach Möglichkeit als eine Funktion des Repositoriums automatisiert erfolgen und keine weiteren Schritte der Datenbereitstellenden erfordern.

5.4 Vermeidung von unnötigen Kopien von Datensätzen

Sekundärdaten (d.h. Transformationen oder Selektionen, die von einem Primärdatensatz abgeleitet wurden; s. Abschn. 2) sollten niemals als ein neuer Primärdatensatz abgelegt werden – dies wäre nicht nur unökonomisch, sondern könnte zu massiven Verzerrungen weiterer Nutzungen führen (etwa wenn beide Datensätze in der Folge als unabhängige Datensätze in eine Meta-Analyse einfließen würden). Stattdessen sollten Nachnutzende immer auf den ursprünglich von den Datenbereitstellenden publizierten Primärdatensatz verweisen und nachvollziehbar dokumentieren, wie ein aus den Originaldaten abgeleiteter Datensatz entstanden ist.

Die unter 5.1. und 5.2 beschriebene Unterscheidung von Typ 1 und Typ 2 Datenveröffentlichungen erlaubt allerdings eine schrittweise Veröffentlichung von (Teil-)

Datensätzen, die möglicherweise auch überlappende Inhalte haben. Um dies mit der erforderlichen Vermeidung von Kopien in Einklang zu bringen, könnten Datenbereitstellende den Gesamtdatensatz (im Sinne von Typ 2) unter einem Nutzungsembargo veröffentlichen (s. Abschn. 5.3) und mithilfe eines reproduzierbaren Skriptes daraus einen Teildatensatz erstellen, welcher als Typ 1 Veröffentlichung zusammen mit einer Publikation sofort frei verfügbar gemacht wird. Im Repositorium sollte klar kenntlich gemacht sein, dass dieser Datensatz von einem anderen Primärdatensatz abgeleitet wurde. Außerdem sollten Primärdatensatz und alle abgeleiteten Teildatensätze im Repositorium gebündelt präsentiert werden, so dass die Abhängigkeit der Datensätze deutlich wird. Grundsätzlich sollte die Anzahl solcher Teildatensätze für jede Datenerhebung so gering wie möglich gehalten und die Beziehung derart „verwandter“ Teil- und Primärdatensätze möglichst transparent kommuniziert werden.

6. Lizenzen und Zugriffsklassen

6.1 Allgemeines

Forscherinnen und Forscher, die Primärdaten produzieren („Datenbereitstellende“), sollten das Recht an der Erstnutzung dieser Daten haben (auch wenn dieses Recht nicht im engeren Sinne juristisch einklagbar sein wird). Dies gilt auch dann, wenn es sich um eine Datenbereitstellung vom Typ 2 handelt (s. Abschn. 5.2) und die Daten zum Zeitpunkt der Bereitstellung noch nicht von den Datenbereitstellenden selbst genutzt worden sind. Datenbereitstellende können sich ihr Erstnutzungsrecht also durch ein befristetes Nutzungsembargo (s. Abschn. 5.3) sichern.

Forscherinnen und Forscher, die Daten anderer nachnutzen, müssen diese Daten adäquat zitieren²⁸. Dafür sollte den Daten im Repositorium ein Zitationshinweis (mit Angabe eines persistenten Identifikators; DOI – s. Abschn. 4) beiliegen. Nachnutzende sind außerdem dazu verpflichtet, die Daten so zu analysieren, dass Rechte der Teilnehmerinnen und Teilnehmer der Originalstudie nicht verletzt werden. Hierfür sind die Nachnutzenden verantwortlich und müssen die von den Datenbereitstellenden im Repositorium angegebenen Nutzungsbedingungen genau beachten.

Im Falle einer Datennachnutzung gelten die gleichen Anforderungen an Transparenz und wissenschaftliche

Sorgfalt wie bei der Erstnutzung. Für eine wissenschaftliche Neuinterpretation der Daten müssen die wissenschaftlichen Standards zum Zeitpunkt der Nachnutzung gelten. Bei der Bewertung von Originalanalysen hingegen sollten im Kontext einer Reanalyse fairerweise diejenigen Standards angelegt werden, die zur Zeit der Originalanalysen galten.

Nachnutzende haben alle zur Verfügung gestellten Dateien (sowie ggf. angelegte Kopien) nach Erlöschen eines vertraglich vereinbarten Nutzungsrechts vollständig zu löschen.

6.2 Zugriffsklassen

Datenschutzrechtliche Gesetze, forschungsethische Erwägungen (vgl. Deutsche Gesellschaft für Psychologie DGPs, 2018) oder begründbare wissenschaftliche Interessen der Datenbereitstellenden können es erforderlich machen, den Zugriff auf einen Datensatz bzw. dessen Nutzung – unabhängig davon, ob es sich um eine Datenbereitstellung vom Typ 1 oder Typ 2 handelt (s. Abschn. 5.1 und 5.2) – zu beschränken. Dabei sollten datenschutzrechtliche und forschungsethische Bedenken stets in einer Risikoanalyse abgewogen werden, in die sowohl das Risiko einer missbräuchlichen Nutzung als auch die Schwere des entstehenden Schadens im Falle einer missbräuchlichen Nutzung einfließen. Beispielsweise können sowohl ein hohes Risiko der Reidentifizierung von Versuchspersonen gepaart mit einem niedrigen erwarteten Schaden, als auch ein niedriges Risiko der Reidentifizierung gepaart mit einem hohen erwarteten Schaden eine starke Zugriffsbeschränkung notwendig machen. Diesbezüglich unterscheiden wir vier Zugriffsklassen, welche im Folgenden mit den Ziffern 0, 1, 2 und 3 gekennzeichnet sind. Die Klassen vereinen jeweils eine technische Zugriffsbeschränkung (Verfügbarmachung) mit Nutzungslizenzen, die die zulässige Verwendung der bezogenen Datensätze regeln. Der Begriff Lizenz bezieht sich hier auf eine vertragliche Vereinbarung, mit der Datenbereitstellende einer zweiten Partei Dinge erlauben können, die ohne diese explizite Vereinbarung gegen Rechte des Datenbereitstellenden verstoßen würden²⁹. An dieser Stelle sei noch einmal darauf hingewiesen, dass Fragen zu Urheberrecht und Nutzungsrechten geklärt sein müssen, bevor ein Datensatz mit einer Lizenz belegt werden kann (s. Abschn. 3.4).

- **Zugriffsklasse 0 („open data“):** Weder der Zugriff noch die Nutzung der Daten sind in irgendeiner Form

²⁸ <https://www.force11.org/datacitation>

²⁹ <https://www.forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/>

restringiert. Die Nutzung kann – abhängig von der jeweiligen Plattform bzw. den Funktionalitäten des Repositoriums – ggf. digital nachverfolgt werden („tracking“). Eventuell werden Datennachnutzende daher gebeten, Angaben zu Identität, Affiliation und/oder Zweck der Nachnutzung zu machen³⁰. Rechtlich kommen hier verbreitete Standardlizenzen z.B. aus der Creative Commons³¹ oder Open Data Commons³²-Familie zum Einsatz.

- **Zugriffsklasse 1 („open data / conditional access“):** Der Zugriff bzw. die Nachnutzung ist an bestimmte, von den Datenbereitstellenden festgelegte Bedingungen geknüpft, mit denen sich Datennachnutzende explizit einverstanden erklären müssen. Bei diesen Bedingungen handelt es sich um Lizenzverträge mit Standardbedingungen (Beispiele s. weiter unten). Nur wenn die Nachnutzenden diesen vertraglich geregelten Bedingungen zustimmen, wird der Zugriff ermöglicht (ohne dass vorher noch einmal eine Bestätigung der Datenbereitstellenden eingeholt wird). Verstöße gegen die Bedingungen können von den Datenbereitstellenden geahndet werden (s. Abschn. 7). Je nach Repositoryum können hier Informationen über die Nachnutzenden (Identität und Affiliation) und/oder den Zweck der Nachnutzung eingeholt und an die Datenbereitstellenden übermittelt werden.
- **Zugriffsklasse 2 („restricted access“):** Der Zugriff bzw. die Nachnutzung ist über die in Zugriffsklasse 1 geltenden Bestimmungen hinaus an weitere, vorher von den Datenbereitstellenden festgelegte Bedingungen geknüpft, mit denen sich Nachnutzende explizit einverstanden erklären müssen. Hier handelt es sich nicht um „Standardfälle“ (wie in Klasse 1), sondern um individualisierte Verträge zwischen Datenbereitstellenden und Nachnutzenden. Der Zugriff wird nur ermöglicht, wenn sich beide Vertragsparteien mit den darin definierten Klauseln vollumfänglich einverstanden erklären. Vertragsverletzungen können von den Datenbereitstellenden geahndet werden (s. Abschn. 7).
- **Zugriffsklasse 3 („secure data“):** Ein Zugriff auf die Daten ist nur unter datenschutzrechtlich gebotenen Restriktionen (z.B. online über einen entsprechend abgesicherten Kanal oder vor Ort in einem Datenzentrum) möglich. Art und Umfang der Nutzung werden durch die Datenbereitstellenden festgelegt und von der entsprechenden Einrichtung, in der sich die Daten befinden, überwacht.

Die Wahl der Zugriffsklasse obliegt den Datenbereitstellenden unter Berücksichtigung etwaiger Vorgaben von Drittmittelgebern. Die Entscheidung sollte begründet und im Zuge der Datenveröffentlichung dokumentiert werden. Im Sinne der Förderung von Offenheit und Transparenz gilt der Grundsatz: So offen wie möglich, so restriktiv wie nötig.

Zugriffsklasse 0 ist in allen Fällen angemessen, in denen (a) personenbezogene Daten (s. Abschn. 3.2) entweder ohnehin nie Bestandteil des Datensatzes waren oder für die Beantwortung der Forschungsfragestellung irrelevant sind (und daher auch aus dem öffentlichen Datensatz entfernt werden können) und (b) es keine aus Sicht der wissenschaftlichen Gemeinschaft nachvollziehbaren (z.B. forschungsethischen) Gründe gibt, derentwegen die Nachnutzung eingeschränkt bzw. begrenzt werden muss.

In **Zugriffsklasse 1** werden standardmäßige Bedingungen für den Zugriff definiert. Das entscheidende Merkmal von Zugriffsklasse 1 ist, dass ein *Standardlizenzvertrag* bei der Nachnutzung abgeschlossen wird. Zwar können verschiedene Repositoryen diese Standardlizenzverträge unterschiedlich ausgestalten, doch der Vorteil dieser Zugriffsklasse für die Wissenschaftsgemeinde entfaltet sich erst dann, wenn eine gewisse Homogenität der entsprechenden Vertragsbedingungen erreicht wird: Wenn allgemeine Standards etabliert sind, können sowohl Datenbereitstellende wie Datennachnutzende effiziente Arbeitsroutinen entwickeln und damit einen übermäßigen Zeitaufwand auf Rechtsangelegenheiten vermeiden. Standardbedingungen für den Zugriff können beispielsweise folgende Aspekte beinhalten:

- die Nutzung eines Datensatzes zu ausschließlich wissenschaftlichen Zwecken („scientific use“; s. Abschn. 6.3),
- im Falle einer Veröffentlichung, die aus der Nachnutzung hervorgeht, die Verpflichtung, auf eine oder mehrere relevante Quellen zu verweisen (z.B. die Originalpublikation),
- das Einholen evtl. erforderlicher Einverständniserklärungen (z.B. von Seiten eines Testverlages, welcher die Rechte an den Normierungsdaten des Tests besitzt),
- die explizite vertragliche Zusage, keinen Reidentifizierungsversuch der Studienteilnehmerinnen und -teilnehmer vorzunehmen, oder
- die Verpflichtung, sich an Nutzungseinschränkungen, die mit den Studienteilnehmenden ausgehandelt wur-

³⁰ Inwiefern diese Informationen dem Datenbereitstellenden übermittelt werden, hängt von dessen Wünschen sowie von den Funktionalitäten des Repositoriums ab.

³¹ <https://creativecommons.org/licenses>

³² <https://opendatacommons.org>

den, zu halten (bspw. einen klinisch bedeutsamen Zufallsbefund im Zuge der Reanalyse zu melden; vgl. Deutsche Gesellschaft für Psychologie DGPs, 2018).

In **Zugriffsklasse 2** können Datenbereitstellende individualisierte Nutzungsbeschränkungen in einem Lizenzvertrag definieren. Diese Beschränkungen müssen nachvollziehbar begründet werden. Die folgenden Beispiele sollen verdeutlichen, was unter „nachvollziehbar“ verstanden werden kann.

- **Beispiel 1:** Bei den bereitgestellten Daten handelt es sich um den Normierungsdatensatz eines Fähigkeits- bzw. Leistungstests, der nur dann valide Ergebnisse liefern kann, wenn die richtigen Antworten nicht öffentlich zugänglich sind. Der Datensatz (der diese Informationen notwendigerweise beinhaltet) kann nachgenutzt werden, aber die Nachnutzenden dürfen diese Information keinesfalls veröffentlichen.
- **Beispiel 2:** Es besteht das Risiko, dass Nachnutzende eine verzerrte und invalide Reanalyse der Originaldaten anstreben könnten, um eine bestimmte Agenda zu verfolgen. Die Vertragsbedingung könnte hier lauten, dass die Nachnutzenden den Datenbereitstellenden im Falle einer geplanten Publikation der reanalysierten Daten immer die Möglichkeit geben, die Reanalyse zu überprüfen bzw. zu kommentieren, und zwar bevor die Befunde der Reanalyse bei einer Zeitschrift o.ä. eingereicht werden. Dieses Vorgehen sollte nur in sehr seltenen Fällen gewählt werden und sollte durch die Datenbereitstellenden öffentlich nachvollziehbar begründet werden.
- **Beispiel 3:** Datenbereitstellende ziehen die Möglichkeit eines Nutzungsembargos in Betracht (s. Abschn. 5.3). Die Zusatzvereinbarung könnte beinhalten, dass die Daten während der Embargoperiode nur zum Zwecke der Reproduktion bereits publizierter Ergebnisse analysiert oder für Metaanalysen genutzt werden dürfen, aber nicht für weitergehende Fragestellungen. Eine solche Einschränkung sollte aber nur dann zulässig sein, wenn das Ende der Embargoperiode klar benannt ist und die Daten danach automatisch in einer weniger restriktiven Zugriffsklasse bereitgestellt werden.

Eine Zugriffsbeschränkung im Sinne von **Zugriffsklasse 3** ist angemessen, wenn die Risikoanalyse ergibt, dass es leicht möglich ist, Individuen zu identifizieren oder wenn den betreffenden Personen durch das Bekanntwerden persönlicher Informationen Schaden entstünde. Darunter könnten hochsensible Daten fallen, wie z. B. Bildung mit Identifikation der Schädelform oder des Gesichts, Genom-Daten, aber auch Umfragedaten zu sensiblen Themen, die im Rahmen klinischer oder organisationspsychologischer Studien erhoben wurden. Nach-

nutzende müssen eine entsprechend restriktive Geheimhaltungs- und Nichtweitergabeerklärung unterzeichnen. Die Anwendung dieser Zugriffsklasse kann in Einzelfällen auch dann angemessen sein, wenn begründbar anzunehmen ist, dass Datensätze für missbräuchliche Zwecke genutzt werden (vgl. Lewandowsky & Bishop, 2016).

Solange keine rechtlichen oder forschungsethischen Einwände bestehen, kann eine festgelegte Zugriffsbeschränkung von den Datenbereitstellenden jederzeit gelockert werden. Eine nachträgliche Restringierung des Zugriffs sollte nur dann erfolgen, wenn sich im Nachhinein herausstellen sollte, dass eine höhere Zugriffsklasse aus rechtlichen oder forschungsethischen Gründen erforderlich ist.

6.3 Scientific Use

Dateien (einschl. Untersuchungsmaterial, Primärdaten, etc.), deren Nutzung ausschließlich auf den Kontext wissenschaftlicher Analysen und Diskussionen begrenzt bleiben sollen, werden als „scientific use files“ (SUF) bezeichnet. Dabei kann die „wissenschaftliche Nutzung“ über den Kreis der Nutzenden oder über den Zweck der Nutzung (oder beides) definiert werden. Bei einer Definition über den *Nutzerkreis* könnte beispielsweise vorausgesetzt werden, dass nur Personen Zugriff haben, die an einer nationalen oder internationalen Forschungseinrichtung tätig sind und/oder eine wissenschaftliche Ausbildung (i. d. R. Promotion) nachweisen können. Da die Automatisierung einer solchen Überprüfung häufig fehleranfällig sein wird, empfehlen wir eine Definition des „scientific use“ über den Nutzerkreis ausdrücklich nicht. Sofern eine Einschränkung des Nutzerkreises gewünscht und gerechtfertigt ist, sollte diese stattdessen über individualisierte Nutzungsverträge in Zugriffsklasse 2 umgesetzt werden.

Bezüglich des *Nutzungszwecks* meint „scientific use“ (a) die Reanalyse eines Datensatzes zur wissenschaftlichen Qualitätssicherung, (b) Veranschaulichungen bzw. Übungen im Kontext universitärer Lehrveranstaltungen, (c) die Überprüfung der Anwendbarkeit eines bestimmten statistischen Verfahrens bzw. der Schätzbarkeit eines statistischen Parameters oder (d) die Nutzung eines Datensatzes zur Beantwortung neuer Forschungsfragestellungen.

Grundsätzlich verbietet sich die Weitergabe von Materialien, die in Zugriffsklasse 1 oder höher eingeordnet sind, an Dritte. Wenn SUF zu Lehrzwecken verwendet werden sollen (etwa weil anhand eines Datensatzes die Verwendung einer bestimmten Analyse verdeutlicht oder die Anwendbarkeit eines bestimmten Modells geprüft werden soll), müssen die Studierenden daher individuell

die entsprechenden Daten direkt aus dem Repositorium beziehen. Zu beachten ist hier, dass auch im Kontext von Lehrveranstaltungen alle Studierenden an die in den jeweiligen Lizenzvereinbarungen definierten Bedingungen gebunden sind.

Die Bereitstellenden von SUF sollten klare Bedingungen für ein Ende der Nutzungsrechte durch die Nachnutzenden definieren: Sobald die Nutzungsrechte erlöschen, etwa wenn der vorgesehene Nutzungszweck erfüllt ist oder aufgegeben wird, sind jegliche Kopien der SUF vom Nachnutzenden zu löschen.

7. Strukturelle Herausforderungen und Empfehlungen

7.1 Würdigung der Datenbereitstellung und Anreizstrukturen

Datenbereitstellung und Datennachnutzung sollten – wie bereits in der Einleitung angesprochen – eine „Win-win-Situation“ für die gesamte wissenschaftliche Gemeinschaft darstellen, zur Qualitätssicherung in der Wissenschaft beitragen und die Generierung neuer, robuster Erkenntnisse sowie methodischer Weiterentwicklungen fördern. Dies alles gelingt aber nur dann, wenn sowohl die Datenbereitstellung als auch die Datennachnutzung im Geiste einer kooperativen und produktiven Maximierung des kollektiven Interesses erfolgen. Dies erfordert nicht nur die Bereitschaft zu Offenheit und Transparenz auf der Seite der Datenbereitstellenden, sondern auch die Bereitschaft der Nachnutzenden, sich an bestimmte Regeln der guten wissenschaftlichen Praxis zu halten. Und schließlich erfordert dies eine angemessene Würdigung des Zeit- und Arbeitsaufwandes, der notwendigerweise mit der Bereitstellung von Daten seitens derjenigen, die sie erhoben haben, einhergeht.

Wie eine solche Würdigung aussehen kann und welche strukturellen Anreize das Wissenschaftssystem hier bereitstellen kann, ist zurzeit Gegenstand intensiver Debatten in unterschiedlichen Fachdisziplinen³³ und ist nicht im engeren Sinne Bestandteil der vorliegenden Empfehlungen zum Datenmanagement. Wir können deshalb an dieser Stelle nur Anregungen und Denkanstöße geben. Eine Form der Würdigung könnte die Möglichkeit sein,

die Bereitstellung von Daten – etwa bei der Evaluation wissenschaftlicher Lebensläufe – ebenso als einen wissenschaftlichen Produktivitätsindikator zu verstehen wie eine Textpublikation. Schon jetzt gibt es entsprechende Zeitschriften, in denen Primärdaten und deren Beschreibung veröffentlicht werden können, wodurch diese entsprechend zitierfähig werden^{34,35}. Wenn Datenbereitstellungen als eigene Kategorie in Lebensläufen gewürdigt werden, kann deren Beitrag auch sichtbar werden. Auch die Vergabe von Preisen (wie der DGPs-Förderpreis zur Qualitätssicherung in der Psychologie oder die *Leamer-Rosenthal Prizes for Open Social Science*³⁶) dürfte insbesondere für Nachwuchswissenschaftlerinnen und Nachwuchswissenschaftler eine angemessene, wenn auch noch nicht hinreichende Würdigung für ihre Bereitschaft zur Datenbereitstellung darstellen.

Wissenschaftliche Fachgesellschaften, Förderinstitutionen, Fachzeitschriften, aber auch Institutionen vor Ort sind bei der Entwicklung und Implementierung entsprechender Anreizstrukturen gefragt. Ziel dieser Anreizstrukturen muss es sein, die Bereitschaft zur Datenbereitstellung und den damit verbundenen Arbeitsaufwand angemessen zu würdigen.

7.2 Koautorenschaften

Sollte aus der Datennachnutzung eine Publikation entstehen, stellt sich die Frage, ob bzw. unter welchen Umständen den Datenbereitstellenden eine Koautorenschaft anzubieten ist. Da – jedenfalls im Moment noch – eine Autorenschaft bei einer publizierten Arbeit zu den am häufigsten herangezogenen Indikatoren wissenschaftlicher Produktivität zählt, könnte man argumentieren, dass eine Koautorenschaft des Datenbereitstellenden bei Publikationen, die aus der Nachnutzung von Daten resultieren, die einfachste und angemessenste Form der Würdigung einer Datenbereitstellung darstellt. Die DGPs-Kommission „Open Science“ hat diese Argumentation intensiv diskutiert und kommt diesbezüglich zu folgender Empfehlung: Wir sind der Auffassung, dass die bloße Bereitstellung von Daten keine Koautorenschaft der Originalautorinnen bzw. -autoren rechtfertigt, da eine Koautorenschaft „...einen genuinen, nachvollziehbaren Beitrag zu dem Inhalt einer wissenschaftlichen Text-, Daten- oder Softwarepublikation“ erfordert (DFG-Leitlinien zur Sicherung guter wissenschaftlicher Praxis vom Septem-

³³ siehe z. B. <https://psyarxiv.com/6btc3/>

³⁴ <https://openpsychologydata.metajnl.com/>

³⁵ <https://www.nature.com/sdata/>

³⁶ <https://www.bitss.org/lr-prizes/>

ber 2019; Leitlinie 14; s. auch Abschnitt 7.3, Absatz 13, Satz b der Berufsethischen Richtlinien der DGPs bzw. des BDP³⁷). Ein solcher Beitrag geht über das bloße Bereitstellen von Daten üblicherweise hinaus und beinhaltet zumindest die substanzielle Mitarbeit am Manuskript selbst. Auch halten wir es für unangemessen, dass Originalautorinnen und -autoren ihre Daten grundsätzlich nur dann bereitstellen, wenn ihnen die Koautorenschaft eines aus der Nachnutzung der Daten ggf. entstehenden Manuskripts in Aussicht gestellt wird.

Andererseits kann es für Nachnutzende in vielen Fällen sinnvoll sein, mit den Datenbereitstellenden Kontakt aufzunehmen, etwa um Missverständnisse, die sich beispielsweise aus der Nicht-Reproduzierbarkeit einer Analyse ergeben, zu klären, oder um die Angemessenheit einer Koautorenschaft der Datenbereitstellenden bei der aus der Reanalyse entstehenden Arbeit zu diskutieren.

Um unterschiedliche Rollen und Beitragsformen verschiedener Koautoren darzustellen, eignet sich die CRediT-Taxonomie³⁸. Eine Einordnung aller Autoren in die 14 verschiedenen CRediT-Beitragsrollen wird aktuell bereits von einer Reihe von Journals standardmäßig verlangt (s. auch McNutt et al., 2018; Holcombe, 2019).

Noch ist nicht abzusehen, wie die standardmäßige Bereitstellung von Daten und deren Reanalyse durch Nachnutzende die wissenschaftsimmanente Sicht auf althergebrachte Produktivitätsindikatoren (wie das Zählen von Publikationen in einem akademischen Lebenslauf) verändern wird. Daran, dass ein solcher Veränderungsbedarf besteht und dass auch die Bereitstellung von Daten (selbst dann, wenn daraus kein selbst verfasstes Manuskript entsteht) als wissenschaftlich produktive Praxis angesehen und angemessen gewürdigt werden muss, besteht jedoch kein Zweifel.

7.3 Vorgehen bei Konfliktfällen

Wann immer in den vorliegenden Empfehlungen auf geltende Gesetze (etwa zum Datenschutz und Urheberrecht; s. Abschn. 3), auf Richtlinien der guten wissenschaftlichen Praxis oder auf die berufsethischen Richtlinien der DGPs bzw. des BDP verwiesen wird, handelt es sich um bindende und prinzipiell sanktionierbare Vorgaben. So sind Verstöße, etwa gegen das Bundesdatenschutzgesetz, zivil- und evtl. sogar strafrechtlich relevant. Unabhängig davon

definieren die DFG-Leitlinien zur guten wissenschaftlichen Praxis aus dem Jahr 2019, wie mit wissenschaftlichem Fehlverhalten umzugehen ist (s. Leitlinien 18 und 19), und die berufsethischen Richtlinien der DGPs bzw. des BDP besagen, dass Verstöße gegen sie durch das Ehrengericht der DGPs³⁹ bzw. des BDP verfolgt und ggf. geahndet werden können. Demgegenüber sind viele der hier formulierten Empfehlungen nicht rechtlich bindend. Umso wichtiger ist es, zentrale und potenziell konfliktträchtige Punkte von vornherein in standardisierten Vereinbarungen (wie in Zugriffsklasse 1 vorgesehen) oder – falls erforderlich – in spezifischen Vereinbarungen (wie in Zugriffsklasse 2 vorgesehen) anzusprechen, um sich entsprechend abzusichern. Verstöße gegen solche definierten Vereinbarungen können als wissenschaftliches Fehlverhalten betrachtet werden; sie sind damit zivilgerichtlich verfolgbar und – falls die beteiligten Parteien Mitglieder der DGPs sind – vor dem Ehrengericht der DGPs verhandelbar (vgl. die Ehrengerichtsordnung⁴⁰). Für den Umgang mit Streitpunkten zwischen Nachnutzenden und Datenbereitstellenden, die nicht eindeutig in die Kategorie „wissenschaftliches Fehlverhalten“ fallen, sondern vielmehr unterschiedliche Sichtweisen darstellen, empfehlen wir die Einrichtung eines (im Idealfall durch Wahl legitimierten) Ombudsgremiums, an das sich die Beteiligten wenden können, sofern sie Mitglieder der DGPs sind.

Fazit

Die Wissenschaft ist im Wandel begriffen. Angetrieben von neuen technischen Entwicklungen, dringlich gemacht durch die sogenannte Replikationskrise, und verbindlich erwartet durch neue Vorgaben von Zeitschriften und Geldgebern, wird die Bereitstellung offener Forschungsdaten auch in der psychologischen Forschung zunehmend zur Norm.

Die hier dargelegten Empfehlungen sollen einen Rahmen bereitstellen, mit dem dieser Schritt hin zu mehr Offenheit und Transparenz den größtmöglichen Nutzen sowohl für Datenbereitstellende als auch Nachnutzende entfalten kann und damit im Idealfall die Qualität der psychologischen Forschung nachhaltig verbessert.

³⁷ <https://www.dgps.de/index.php?id=85#c2001838>

³⁸ <https://casrai.org/credit>

³⁹ <https://www.dgps.de/index.php?id=78#c281>

⁴⁰ https://www.dgps.de/fileadmin/documents/Fachgruppen/DGPs_Ehrengerichtsordnung_2017.pdf

Preprints

Die deutsche Originalversion dieses Texts kann hier abgerufen werden: <https://doi.org/10.31234/osf.io/hcxtm>. Die englische Übersetzung, die von Lisa Trierweiler (ZPID Trier) in Zusammenarbeit mit der DGPs-Kommission „Open Science“ erarbeitet wurde, kann hier abgerufen werden: <https://doi.org/10.31234/osf.io/24ncs>.

Literatur

- Abele-Brehm, A., Gollwitzer, M., Steinberg, U. & Schönbrodt, F. (2019). Attitudes towards Open Science and public data sharing: A survey among members of the German Psychological Society. *Social Psychology*, 50, 252–260. <https://doi.org/10.1027/1864-9335/a000384>
- Deutsche Gesellschaft für Psychologie DGPs (Hrsg.). (2018). *Ethisches Handeln in der psychologischen Forschung*. Göttingen: Hogrefe.
- Gollwitzer, M., Schönbrodt, F. D., Steinberg, U. & Abele-Brehm (2018). Die Datenmanagement-Empfehlungen der DGPs: Ein Zwischenstand. *Psychologische Rundschau*, 69, 366–373. <https://doi.org/10.1026/0033-3042/a000415>
- Guibault, L. & Wiebe, A. (Hrsg.). (2013). *Safe to be open*. Göttingen: Universitätsverlag Göttingen.
- Hillegeist, T. (2012). *Rechtliche Probleme bei der elektronischen Langzeitarchivierung wissenschaftlicher Primärdaten*. Göttingen: Universitätsverlag Göttingen.
- Holcombe, A. O. (2019). Contributorship, Not Authorship: Use CRediT to Indicate Who Did What. *Publications*, 7 (3), 48. <https://doi.org/10.3390/publications7030048>
- Hox, J. J. & Boeije, H. R. (2005). Data Collection, Primary vs. Secondary. In K. Kempf-Leonard (Hrsg.), *Encyclopedia of Social Measurement* (Vol. 1, S. 593–599). New York: Elsevier.
- Lewandowsky, S. & Bishop, D. (2016). Research integrity: Don't let transparency damage science. *Nature*, 529 (7587), 459–461. <http://doi.org/10.1038/529459a>
- Martin, N., Friedewald, M., Schierung, I., Mester, B. A. & Hallinan, D. (Hrsg.). (2020). *Die Datenschutz-Folgenabschätzung nach Art. 35 DSGVO: Ein Handbuch für die Praxis*. Karlsruhe: Fraunhofer Verlag.
- McNutt, M. K., Bradford, M., Drazen, J. M., Hanson, B., Howard, B., Jamieson, K. H., ... & Verma, I. M. (2018). Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proceedings of the National Academy of Sciences*, 115 (11), 2557–2560. <https://doi.org/10.1073/pnas.1715374115>
- Metschke, R. & Wellbrock, R. (2002). *Datenschutz in Wissenschaft und Forschung*. Berlin: Berliner Beauftragter für Datenschutz und Informationsfreiheit.
- Rat für Sozial- und Wirtschaftsdaten RatSWD. (2018). Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften – Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte (2. Aufl.). RatSWD Output 3 (5) [Online-Dokument]. <https://doi.org/10.17620/02671.7>
- Rat für Sozial- und Wirtschaftsdaten RatSWD. (2020). *Datenerhebung mit neuer Informationstechnologie. Empfehlungen zu Datenqualität und -management, Forschungsethik und Datenschutz*. RatSWD Output 6 (6) [Online-Dokument]. <https://doi.org/10.17620/02671.47>
- Schönbrodt, F. D., Gollwitzer, M. & Abele-Brehm, A. (2017). Der Umgang mit Forschungsdaten im Fach Psychologie. *Psychologische Rundschau*, 68, 20–35. <https://doi.org/10.1026/0033-3042/a000341>
- Sim, I., Stebbins, M., Bierer, B. E., Butte, A. J., Drazen, J., Dzau, V., ... & Li, R. (2020). Time for NIH to lead on data sharing. *Science*, 367 (6484), 1308–1309. <https://doi.org/10.1126/science.aba4456>
- Spindler, G. & Hillegeist, T. (2011). Rechtliche Probleme der elektronischen Langzeitarchivierung von Forschungsdaten. In S. Büttner, H.-C. Hobohm & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement*. Bad Honnef: Bock und Herchen.
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Prof. Dr. Mario Gollwitzer

Ludwig-Maximilians-Universität München
Department Psychologie
Lehrstuhl für Sozialpsychologie
Leopoldstraße 13
80802 München
mario.gollwitzer@lmu.de

<https://doi.org/10.1026/0033-3042/a000514>