

## **Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft - AG Forschungsdaten im Auftrag des Vorstands der DGPuK**

**Christina Peter, Johannes Breuer, Philipp K. Masur, Michael  
Scharrow, Christian Schwarzenegger**

### **Angaben zur Veröffentlichung / Publication details:**

Peter, Christina, Johannes Breuer, Philipp K. Masur, Michael Scharrow, and Christian Schwarzenegger. 2020. "Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft - AG Forschungsdaten im Auftrag des Vorstands der DGPuK." *Studies in Communication and Media* 9 (4): 599-626. <https://doi.org/10.5771/2192-4007-2020-4-599>.

## INVITED PAPER

### **Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft**

AG Forschungsdaten im Auftrag des Vorstands der DGPUK

**Guidelines for handling research data in communication science**  
Working group research data on behalf of the management board  
of the German Communication Association (DGPuK)

*Christina Peter, Johannes Breuer, Philipp K. Masur,  
Michael Scharkow & Christian Schwarzenegger*

**Christina Peter (Dr.)**, Institut für Kommunikationswissenschaft und Medienforschung, Ludwig-Maximilians-Universität München, Oettingenstr. 67, 80538 München, Germany. Contact: christina.peter(at)ifkw.lmu.de.

**Johannes Breuer (Dr.)**, Datenarchiv für Sozialwissenschaften, GESIS – Leibniz Institut für Sozialwissenschaften, Unter Sachsenhausen 6-8, 50667 Köln, Germany. Contact: johannes.breuer(at)gesis.org.

**Philipp K. Masur (Dr.)**, Department of Communication Science, Vrije Universiteit Amsterdam, De Boelelaan 1081 / 1081 HV Amsterdam, Netherlands. Contact: p.k.masur(at)vu.nl.

**Michael Scharkow (Prof. Dr.)**, Institut für Publizistik, Johannes Gutenberg-Universität Mainz, Jakob-Welder-Weg 12, 55128 Mainz, Germany. Contact: scharkow(at)uni-mainz.de.

**Christian Schwarzenegger (Dr.)**, Philosophisch-Sozialwissenschaftlichen Fakultät, Universität Augsburg, Universitätsstraße 10, 86159 Augsburg, Germany. Contact: christian.schwarzenegger(at)phil.uni-augsburg.de.

## INVITED PAPER

### **Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft**

AG Forschungsdaten im Auftrag des Vorstands der DGPUK

#### **Guidelines for handling research data in communication science**

Working group research data on behalf of the management board of the German Communication Association (DGPUK)

*Christina Peter, Johannes Breuer, Philipp K. Masur,  
Michael Scharkow & Christian Schwarzenegger*

**Anmerkung:** Zusätzlich danken wir Olaf Jandura, Anna Schnauber-Stockmann, Jörg Hagenah und Corinna Lühje für ihr Engagement in Rahmen der AG Forschungsdaten.

**Zusammenfassung:** Forschungsdaten sind das primäre Produkt empirischer Studien und die zentrale Grundlage wissenschaftlicher Erkenntnis. Ihre langfristige Sicherung und Bereitstellung zur intersubjektiven Nachvollziehbarkeit und damit zur Qualitätssicherung sollte daher in allen wissenschaftlichen Disziplinen unterstützt werden. Entsprechend hat die Deutsche Forschungsgemeinschaft 2016 die sozialwissenschaftlichen Fachgesellschaften in Deutschland aufgefordert, sich mit dem Thema zu befassen und konkrete Leitlinien zum Umgang mit Forschungsdaten zu entwickeln. Die vorliegenden Leitlinien sind dabei aus einer Arbeitsgruppe der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft entstanden. Sie beziehen sich auf alle in der Kommunikationswissenschaft generierten Datentypen und reflektieren dabei die Bedeutung der Bereitstellung von Forschungsdaten, aber auch Herausforderungen und Einschränkungen. Daraus werden Empfehlungen für verschiedene Zielgruppen abgeleitet, um die Bereitstellung von Forschungsdaten in der Kommunikationswissenschaft stärker zu fördern.

**Schlagwörter:** Forschungsdaten, Datenmanagement, Transparenz, Open Science, Replikationen

**Abstract:** Research data are the primary product of empirical studies and essential for generating scientific knowledge. Their accessibility and long-term preservation should therefore be encouraged by all scientific disciplines. Accordingly, the German Research Foundation asked the social sciences associations in Germany to address the topic and develop guidelines for the handling of research data in their discipline. A working group of the German Communication Association (DGPUK) developed the present guidelines. They refer to all data types generated in communication science and reflect the importance of making research data available, while also taking into account challenges and limitations. Recommendations for different target groups are provided to promote the availability of research data in communication science.

**Keywords:** research data, data management, open science, transparency, replication

## Empfehlungen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft

Die Deutsche Forschungsgemeinschaft (DFG) hat im Jahr 2016 die sozialwissenschaftlichen Fachgesellschaften in Deutschland aufgefordert, sich mit dem Thema Forschungsdaten zu befassen und konkrete Leitlinien zum Umgang mit diesen zu entwickeln. Ziel ist es, durch die langfristige Sicherung und Bereitstellung von Forschungsdaten zur intersubjektiven Nachvollziehbarkeit und damit zur Qualitätssicherung in den Sozialwissenschaften beizutragen (vgl. auch DFG, 2015). Diese Bemühungen sind mittlerweile in der DFG-Initiative Nationale Forschungsdateninfrastruktur (NFDI) gebündelt, innerhalb derer der Umgang mit und der Zugang zu Forschungsdaten weiter vorangetrieben werden soll (DFG, 2019).

Als erste Fachgesellschaft hat die Deutsche Gesellschaft für Psychologie (DGPs) solche Leitlinien verabschiedet (Schönbrodt et al., 2016). Im Januar 2019 erfolgte die Veröffentlichung einer entsprechenden Stellungnahme durch die Deutsche Gesellschaft für Soziologie (DGS, 2019). Auf diese hat die Akademie für Soziologie (AS) eine Replik verfasst (AS, 2019a) und im September 2019 dann eigene Richtlinien formuliert (AS, 2019b). Das folgende Dokument schließt in weiten Teilen an die genannten Leitlinien an, berücksichtigt jedoch weiterhin Besonderheiten kommunikationswissenschaftlicher Forschung und die daraus resultierenden Herausforderungen für den Umgang mit Forschungsdaten, insbesondere die starke Verbreitung inhaltsanalytischer Studien sowie generell die Generierung von qualitativen und quantitativen Forschungsdaten im Fach.<sup>1</sup>

### 1. Die Bedeutung von Forschungsdaten und ihrer Verfügbarkeit für die Kommunikationswissenschaft

Forschungsdaten sind das primäre Produkt empirischer Studien und die zentrale Grundlage wissenschaftlicher Erkenntnis. Wie in der Wissenschaft allgemein, sind auch in der Kommunikationswissenschaft die Einhaltung methodischer (Qualitäts-)Standards in der Erhebung und Auswertung von Daten die notwendige Voraussetzung für einen Erkenntniszuwachs. Dies gilt grundlegend sowohl für qualitative wie auch quantitative Forschung, wenngleich sich die jeweiligen Gütekriterien sowie ihre spezifische Gewichtung zwischen diesen beiden Forschungssträngen unterscheiden.

Ein weiteres Kriterium, das neben der Beachtung von Standards und Gütekriterien nicht nur das Vertrauen in die Ergebnisse empirischer Forschung erhöht, sondern auch wesentlich dazu beitragen kann, dass effizienter kumulativer Erkenntniszuwachs und Selbstkorrektur möglich sind, ist die Transparenz der Forschung. Allgemein gesprochen dient Transparenz im Forschungsprozess dazu, die Nachvollziehbarkeit und Verifizierbarkeit von Befunden zu gewährleisten. Ein wichtiges

1 Die Leitlinien wurden von der AG Forschungsdaten formuliert, in mehreren Überarbeitungsschleifen mit Feedback von Vorstand und Fachgruppensprecherinnen überarbeitet und am 11. März 2020 auf der DGPK-Jahrestagung in München von den DGPK-Mitgliedern in Rahmen der Vollversammlung einstimmig verabschiedet. Der vorliegende Beitrag stellt diese finale, abgestimmte Version dar und hat entsprechend kein klassisches Zeitschriften-Peer-Review durchlaufen.

Instrument zur Gewährleistung von Transparenz ist das Verfügbarmachen von Forschungsdaten (sowie des genutzten Materials und der Analyseskripte). Sind diese für andere zugänglich und gut dokumentiert, kann weitergehend untersucht werden, wie robust, reproduzierbar, replizierbar und generalisierbar die auf diesen Daten basierenden Befunde sind bzw. wie plausibel und nachvollziehbar interpretative Schlüsse aus dem Material entwickelt worden sind.

Robust sind Befunde, wenn man mit verschiedenen Auswertungsmethoden für dieselben Daten zu den gleichen Schlüssen kommt. Reproduzierbarkeit bedeutet, dass sich die Ergebnisse mit den gleichen Daten und Auswertungsmethoden erneut hervorbringen lassen, während mit Replizierbarkeit gemeint ist, dass man die Studie wiederholen kann (d.h. neue Daten erhebt) und mit den gleichen Auswertungsmethoden zu gleichen Ergebnissen kommt. Und als generalisierbar kann man Befunde bezeichnen, wenn man mit anderen Methoden, Daten und ggfs. auch Auswertungsmethoden zu den gleichen Schlüssen kommt. Tabelle 1 stellt diese Kriterien noch einmal anschaulich gegenüber.

**Tabelle 1. Gütekriterien für Forschungsbefunde**

		Erhebungsmethoden			
		Gleich		Unterschiedlich	
		Stichprobendaten		Stichprobendaten	
		Gleich	Unterschiedlich	Gleich	Unterschiedlich
Auswertungsmethoden	Gleich	Reproduzierbar	Replizierbar	–	–
	Unterschiedlich	Robust	–	–	Generalisierbar

Anmerkung: Basierend auf LeBel, McCarthy, Earp, Elson & Vanpaemel, 2018.

Oftmals wird argumentiert, dass Reproduzierbarkeit und Generalisierbarkeit für Teile der kommunikationswissenschaftlichen Forschung nicht erreichbar und daher vielleicht auch nicht erstrebenswert sind. Beispiele hierfür könnten u.a. Inhaltsanalysen von Medien- und Kommunikationsinhalten in bestimmten Zeiträumen und Regionen oder auch Studien zu bestimmten Ereignissen (z.B. Bundestagswahlen) sein. Angesichts dessen, dass sich Medieninhalte, aber auch -formate und -techniken rasant ändern können, sollten wir nicht erwarten, dass Ergebnisse entsprechender Studien über die Zeit stabil bleiben. Ausgangspunkt dieser Diskussion sind die unterschiedlichen Ziele, die empirische Forschung, speziell auch in der Kommunikationswissenschaft, verfolgt: Neben der reinen Beschreibung von Phänomenen (z. B. der deskriptiven Analyse repräsentativer Daten), unterscheiden wir vorrangig konfirmatorische und explorative Forschung. Während der konfirmatorische Ansatz versucht, a priori formulierte Hypothesen anhand von empirischen Daten zu testen, geht es beim explorativen Ansatz darum, Hypothesen aus den Daten zu generieren. Das oben beschriebene Ideal der

Replizier- und Generalisierbarkeit bezieht sich dabei vorrangig auf den konfirmatorischen Ansatz. Andererseits zielt ein Teil der kommunikationswissenschaftlichen Forschung auf das Testen von Theorien ab, die häufig nicht zeit- oder kontextspezifisch formuliert wurden und daher allgemeine und universelle Prinzipien der Erstellung, Nutzung, Rezeption und Wirkung von Kommunikationsinhalten postulieren. Hier ist das Ziel durchaus Replizier- und Generalisierbarkeit.

Es ist daher naheliegend, dass die in jüngster Zeit vieldiskutierte Replikationskrise in der Psychologie (z.B. Pashler & Harris, 2012) und die Probleme, die damit einhergehend thematisiert werden, auch die (konfirmatorische) Kommunikationswissenschaft betreffen können. Zum einen betreffen viele der im Zuge der Replikationskrise diskutierten Probleme sozialwissenschaftliche oder sogar empirische Forschung allgemein und dies gilt gleichsam für die meisten der vorgeschlagenen Lösungsansätze. Zum anderen ist es auch für die Kommunikationswissenschaft wichtig, sich darüber zu verständigen, welche Kriterien zur Bewertung von Forschungsergebnissen für sie gelten können und wie diese definiert sein sollen. Dabei sind Abweichungen von Normen in anderen Disziplinen durchaus möglich, jedoch müssen diese gut begründet und transparent sein. So kann Generalisierbarkeit im Sinne von Übertragbarkeit (z.B. auf andere Medien, Inhalte, Ereignisse oder Zeiträume) eher als relevante Forschungsfrage (z.B. "Inwiefern lassen sich die Befunde zur Berichterstattung zur Bundestagswahl 2017 auf die zur Bundestagswahl 2021 übertragen"?) statt als Evaluationskriterium für die Güte bestimmter Befunde aufgefasst werden. Im Hinblick auf Replikationen sind für einige kommunikationswissenschaftliche Studien lediglich konzeptionelle Replikationen möglich, während z.B. in der laborexperimentellen Forschung die direkte Replikation (d.h. mit den exakt gleichen Methoden) maßgeblich für die Bewertung früherer Befunde ist.

Letztlich ist das Verfügbarmachen von kommunikationswissenschaftlichen Forschungsdaten jedoch unabhängig von der Art der durchgeführten Forschung. Obwohl Transparenz in der konfirmatorischen Forschung besonders wichtig ist, um die Robustheit, Reproduzierbarkeit, Replizierbarkeit und Generalisierbarkeit langfristig zu gewährleisten, so gilt auch für deskriptive Analysen und explorative Forschung, dass erst die Verfügbarkeit der Daten das vollständige Nachvollziehen, wie auf Basis der Daten die Ergebnisse entstanden sind, ermöglicht. Das Teilen von Erhebungs- und Analysemethoden erlaubt dann auch die konzeptuelle Replikation, die dann die Frage beantworten kann, ob ursprüngliche Befunde kontext- oder zeitspezifisch eingeordnet werden müssen.

Zusätzlich dazu erhöht das Teilen von Forschungsdaten auch die Effizienz der Forschung, da mit denselben Daten idealerweise unterschiedliche Fragen beantwortet werden können (durch zusätzliche bzw. weitergehende Analysen). Dies ist nicht nur forschungsökonomisch, sondern hat auch eine ethische Komponente, da es im Fall von Daten, die durch Studienteilnehmer\*innen generiert wurden (z.B. in Befragungen, Beobachtungsstudien oder Experimenten), den Nutzen der Daten erhöht und damit den Aufwand der Teilnehmer\*innen umso mehr rechtfertigt. Das kann gerade auch für qualitativ erhobene Daten interessant sein, die in der Datengewinnung oftmals mit erhöhtem zeitlichen Aufwand sowohl seitens der Forschenden wie auch seitens der Befragten verbunden sind. Weiterhin ist ein ökonomischer

bzw. effizienter Umgang mit Ressourcen in der Wissenschaft erstrebenswert, da diese in den meisten Fällen – zumindest in der kommunikationswissenschaftlichen Forschung in Deutschland – aus öffentlichen Geldern finanziert werden. Nicht zuletzt sollten im Fall von öffentlich finanzierten Projekten die Ergebnisse dieser Forschung – soweit wie möglich – auch der allgemeinen Öffentlichkeit zugänglich gemacht werden. Die Akademie für Soziologie verweist hier auf wissenschaftliche Daten als kollektives Gut (AS, 2019a, vgl. auch DFG, 2009).

Neben den (wissenschafts-)ethischen Gründen für die Verfügbarmachung von Forschungsdaten bringt ein strukturiertes, von Beginn des Forschungsprozesses an praktiziertes Forschungsdatenmanagement auch ganz praktische Vorteile für die Wissenschaftler\*innen selbst. Da Forscher\*innen oftmals in mehrere Studien oder Projekte involviert sind und zwischen der Konzeption und Durchführung einer Studie und der Publikation der Ergebnisse in der Regel einige Zeit vergeht, kann es für die Datenproduzent\*innen selbst schwierig sein, nach mehreren Wochen oder Monaten alle Entscheidungen und Schritte im Prozess der Erhebung und Auswertung noch nachzuvollziehen oder zu reproduzieren. Daher ist es vor allem auch für die Forscher\*innen selbst (und ihre Kooperationspartner\*innen) wichtig, die Erhebung und Analyse der Daten gut zu dokumentieren. Eine genaue Dokumentation mit dem Ziel, die Daten auch für andere nachvollziehbar und nutzbar zu machen, erleichtert es entsprechend auch den Forscher\*innen, die diese erhoben haben, damit zu arbeiten. Eine solche Dokumentation steht auch nicht im Widerspruch zur Anonymisierung von Daten oder zum diskreten Umgang mit hochsensiblen Material wie etwa aus ethnographischen Feldnotizen, sondern kann gerade eine Bedingung eines effizienten Umgangs mit schützenswerten Daten sein und etwa eine abgestufte und teilweise Bereitstellung von Daten erst ermöglichen.

Das Bereitstellen von Daten für andere Wissenschaftler\*innen erhöht zudem die Sichtbarkeit der eigenen Forschung und kann neue Möglichkeiten für Kooperationen oder Publikationen eröffnen (McKiernan et al., 2016). Eine Untersuchung von Piwowar und Vision (2013) zeigte in diesem Zusammenhang, dass das Teilen von Forschungsdaten zu einer höheren Zahl von Zitationen für die damit verbundenen Publikationen führt.

Das Teilen von Forschungsdaten ist letztendlich auch eine Antwort auf das große öffentliche Interesse an vielen Fragen und Erkenntnissen der Kommunikationswissenschaft. Ein Fach, das in der öffentlichen Wahrnehmung eine Rolle spielt und spielen will, sollte auch das Interesse haben, dass seine Erkenntnisse einer maximal kritischen Begutachtung durch eine maximale Anzahl von Personen standhalten können. Wenn das Fach möchte, dass seine Befunde in der (medialen) Öffentlichkeit als vertrauenswürdig und haltbar wahrgenommen werden, sollten die Daten, die diesen zugrunde liegen, auch der Öffentlichkeit (z.B. interessierten Fach- und Datenjournalist\*innen) zugänglich sein.

## 2. Herausforderungen bei der Bereitstellung von kommunikationswissenschaftlichen Forschungsdaten

Während der Nutzen der Bereitstellung kommunikationswissenschaftlicher Forschungsdaten somit auf der Hand liegen sollte, stellen sich bei der konkreten Ausgestaltung einige Fragen bzw. Herausforderungen. Neben rechtlichen Aspekten betrifft das unter anderem den Umgang mit Daten, die schwierig oder gar nicht zu anonymisieren sind und für die entsprechend größere Einschränkungen gelten (siehe auch Abschnitt 3.2). Gerade Forschungsbereiche wie die Unternehmens- oder Gesundheitskommunikation können äußerst sensible Daten generieren, die oft weder vollständig anonymisiert werden können (oder durch eine versuchte Anonymisierung entkontextualisiert und damit unbrauchbar würden), aber auch nicht (in Gänze) weitergeben werden dürfen. Ähnliches gilt auch für andere Forschungsgegenstände und Fragestellungen, z.B. biografische Interviews, Befragungen bzw. Interviews mit Zielgruppen wie etwa verfolgten Journalist\*innen oder Aktivist\*innen oder auch inhaltsanalytische Untersuchungen sensibler Dokumente bzw. Aufnahmen. In solchen Fällen muss dem Datenschutz eine höhere Priorität eingeräumt werden als dem Ideal der größtmöglichen Transparenz in der Forschung.

Hierzu wurden bereits in der Stellungnahme der DGS (2019) einige zentrale Punkte festgehalten. Zu diesen gehört unter anderem, dass nicht alle Forschungsdaten (in gleichem Umfang) einer Nachnutzung zugänglich gemacht werden können und dass sich die Auswahl von Forschungsgegenständen (und vor allem auch die Verteilung von Drittmitteln) entsprechend nicht an der Möglichkeit einer solchen Nachnutzung ausrichten sollte, da sonst die Gefahr besteht, dass Fragestellungen und darauf aufbauende Forschungsprojekte, die sensible Daten generieren, marginalisiert werden.

Auch weist die DGS darauf hin, dass heterogene Forschungsansätze und die damit verbundenen Typen von Forschungsdaten einer differenzierten Auseinandersetzung im Rahmen der Diskussion um die Bereitstellung von Forschungsdaten bedürfen. Dies gilt im Besonderen auch für die Kommunikationswissenschaft, da die Bandbreite an empirischer Forschung in unserem Fach von vollständig standardisierter, quantitativer (Experimental-)Forschung bis hin zu offenen qualitativen Ansätzen reicht. Hinzu kommt als genuin kommunikationswissenschaftliche Methode die Inhaltsanalyse und damit die Generierung von – qualitativen wie quantitativen – inhaltsanalytischen Forschungsdaten. Somit generiert die Kommunikationswissenschaft im Vergleich zu anderen Sozialwissenschaften heterogene Datenbestände, denen bei der Frage nach einer Bereitstellung und Nachnutzung der Daten Rechnung getragen werden muss.

Deshalb sei an dieser Stelle explizit erwähnt, dass sich diese Richtlinien und damit die Empfehlung zur Bereitstellung von Forschungsdaten zwar generell auf alle in der Kommunikationswissenschaft erzeugten Forschungsdaten gleichermaßen beziehen, aber auch berücksichtigt wird, dass nicht alle in der Kommunikationswissenschaft generierten Forschungsdaten bereitgestellt werden können und dies in den entsprechenden Fällen auch respektiert werden muss.

### 3. Leitlinien zur Bereitstellung und Nachnutzung von kommunikationswissenschaftlichen Forschungsdaten

Mit den hier skizzierten Leitlinien für den Umgang mit sowie die Bereitstellung und Nachnutzung von Forschungsdaten schließen wir uns an die bereits erarbeiteten Vorschläge der Fachgesellschaften der Psychologie (Schönbrodt et al., 2016) sowie der Soziologie (AS, 2019b; DGS, 2019) an. Als weitere allgemeine Grundlage verweisen wir auf die FAIR-Prinzipien zum Umgang mit Forschungsdaten (Wilkinson et al., 2016), auf die sich auch die Richtlinien der Akademie für Soziologie (AS, 2019b) konkret beziehen. Gemäß dieser sollten Daten auffindbar, zugänglich, interoperabel und wiederverwertbar sein. Aufbauend auf diesen Grundlagen und in Ergänzung dazu skizzieren wir im Folgenden die von uns für das Fach Kommunikationswissenschaft als wichtig erachteten Prinzipien in der Bereitstellung und der Nachnutzung von Forschungsdaten.

#### 3.1 Art der Daten

Wenn wir im Folgenden von Daten oder Forschungsdaten sprechen, meinen wir in Anlehnung an die Empfehlung der DGPs (Schönbrodt et al., 2016) so genannte Primärdaten. Dabei handelt es sich um die erste Übertragung der Rohdaten (z.B. Papierfragebögen, Audioaufzeichnungen von Leitfadeninterviews, Zeitschriftenartikel) in ein digitales Format. Konkrete Beispiele hierfür sind der numerische Code „1“ für eine Ja-Antwort in einer quantitativen Befragung, der durch eine \*n Codier\* in festgehaltene Code für die Art des Mediums, anonymisierte Transkripte von qualitativen Interviews, durch (Computer-)Simulationen entstandene Daten, anonymisierte Log-Daten getrackter Internetnutzung oder die numerisch codierte Erfassung einer physiologischen Messung. Primärdaten sind entsprechend unbearbeitete, nicht aggregierte, aber bereits anonymisierte Forschungsdaten.

Eine Bereitstellung von Rohdaten (also die ausgefüllten Fragebögen, das codierte Material, die Gesprächsaufzeichnungen etc.) dürfte in vielen Fällen schwierig sein: Diese sind oftmals schwer zu anonymisieren oder extrem umfangreich und bedeuten nicht zuletzt auch aus Sicht der Nachnutzenden mehr Aufwand. Insbesondere bei qualitativen Interviews können Rohdaten – z. B. die Audioaufzeichnungen der Gespräche – aus datenschutzrechtlicher Sicht häufig nicht geteilt werden, da die Aufzeichnung der Stimme eine De-Anonymisierung ermöglichen würde.

Während etwa bei Befragungs- oder Beobachtungsstudien der Schutz von Personendaten im Vordergrund steht, sind vor allem bei Inhaltsanalysen auch urheberrechtliche Aspekte relevant. Medieninhalte, die in solchen Studien codiert wurden, dürfen in den meisten Fällen aus urheberrechtlichen Gründen nicht weitergegeben werden. Auch Social-Media-Daten können häufig nicht ohne Weiteres geteilt werden, da die Terms of Service (ToS) der Plattformen die Weitergabe von Daten nicht erlauben oder zumindest einschränken. Diese Einschränkung betrifft sowohl Daten, die über Application Programming Interfaces (APIs) gewonnen wurden, als auch den Kauf von Daten über Dienstleister oder Daten aus direkten Forschungs Kooperationen mit den Plattformen. Eine umfassende Diskussion der Herausforderungen beim Teilen von Social-Media-Daten kann das

vorliegende Dokument nicht leisten, es gibt aber bereits einige Publikationen mit Beispielen und Empfehlungen hierzu (Bishop & Gray, 2017; Kinder-Kurlanda et al., 2017; Mannheimer & Hull, 2017; Thomson, 2016; Williams et al., 2017).

In den Fällen, in denen entsprechend selbst die Primärdaten aus verschiedenen Gründen nicht geteilt werden können, sollte dennoch in Betracht gezogen werden, an ihrer Stelle zumindest aggregierte (z. B. Kovarianz-Matrizen) oder synthetische Daten (Quintana, 2019) bereitzustellen. Dies ist vor allem auch für den Fall von personenbezogenen, nutzergenerierten Inhalten aus der Online-Kommunikation (z.B. Social-Media-Daten) eine interessante Option, wenn diese etwa aufgrund von Nutzungsbedingungen bestimmter Plattformen nicht geteilt werden können. Diese erlaubt zwar keine direkte Nachnutzung, kann aber zum Reproduzieren der Ergebnisse sowie in Teilen sogar für Meta-Analysen genutzt werden.

Generell sollte für das Teilen von Forschungsdaten der Leitsatz gelten: “As open as possible, as restrictive as necessary”, und der Aufwand sollte sich am erwarteten Nachnutzungswert orientieren. Der Aufwand ergibt sich speziell durch die Aufbereitung bzw. Kontextualisierung der Datensätze, aber auch die durch Kommunikation mit Nachnutzenden. Hier müssen Rahmenbedingungen geschaffen werden, die dies im Rahmen von drittmittelgeförderten Forschungsprojekten finanzieren (wie es z.B. bei der DFG bereits vorgesehen ist) bzw. bei nicht drittfINANZIERTEN Projekten incentivieren. Wie weiter oben ausgeführt, kann ein von Beginn des Projektes an gut geplantes Forschungsdatenmanagement den Aufwand für die spätere Bereitstellung erheblich reduzieren (vgl. Klein et al., 2018).

### 3.2 Umfang und Zeitpunkt der Veröffentlichung

Zentral bei der Bereitstellung von Forschungsdaten sind Aspekte des Datenschutzes bzw. des Urheberrechts. Dieser Punkt wurde bereits in den DGPs-Richtlinien (Schönbrodt et al., 2016) ausführlich thematisiert. Grundlegend wichtig ist zunächst, dass die Daten vor dem Teilen so weit anonymisiert werden müssen, dass die Identifikation einzelner Personen nicht (ohne unverhältnismäßigen Aufwand) möglich ist.<sup>2</sup> Grundsätzlich sollte immer eine Einwilligung der Befragten eingeholt werden – bei der Formulierung dieses Informed Consent ist zusätzlich darauf zu achten, ob personenbezogene Daten erhoben werden (siehe Empfehlung bei Schönbrodt et al., 2016). Befragte bzw. Studienteilnehmende sollten entsprechend bereits bei der Rekrutierung darüber informiert werden, dass die von ihnen generierten Antworten gespeichert und Dritten zugänglich gemacht werden, auch wenn eine Anonymisierung möglich ist. Ist das Einholen einer informierten Einwilligung nicht möglich – etwa bei der Sammlung von Social-Media-Daten über Application Programming Interfaces (APIs) oder dem Kauf solcher Daten bei Drittanbietern – sollten die Daten so weit anonymisiert oder aggregiert werden, dass keine direkte Identifikation von Personen mehr möglich ist (sofern es sich nicht um Personen des öffentlichen Lebens handelt).

2 Zur Diskussion über verschiedene Formen der Anonymisierung und der entsprechenden Rechtsprechung siehe Watteler und Kinder-Kurlanda (2015).

Die zur Verfügung gestellten Daten sollten so aufbereitet sein, dass eine Nachnutzung sinnvoll möglich ist. Die veröffentlichten Datensätze sollten alle erhobenen Variablen/Fragen/Kategorien enthalten, auch wenn diese nicht Teil einer Publikation zum Datensatz sind (dies ist in der Publikation auch zu vermerken), da gerade solche Daten für eine Nachnutzung interessant sein können. Zudem sollte zur besseren Nachvollziehbarkeit das jeweilige Mess- oder Erhebungsinstrument (also Fragebogen, Leitfaden, Codebuch, Beobachtungsprotokoll, Stimulusmaterial etc.) mitveröffentlicht werden. Die Daten sollten in der Sprache der Erhebung möglichst so beschriftet bzw. gestaltet sein, dass die Zuordnung zu den entsprechenden Stellen im Messinstrument ersichtlich ist. Ist eine Veröffentlichung des Mess- bzw. Erhebungsinstruments nicht möglich, so sollten die Daten selbst möglichst nachvollziehbar beschriftet sein. Sollten Daten aus dem Originaldatensatz bzw. Originalmaterial gelöscht worden sein (z.B. einzelne Fälle, da das Einverständnis der Person fehlte oder Variablen und Elemente, die aus Anonymisierungsgründen entfernt werden mussten), so ist dies in einem gesonderten Dokument zur Beschreibung des Datensatzes festzuhalten.

Der Zeitpunkt der Veröffentlichung der Daten kann abhängig vom Umfang sowie von der Finanzierung der Daten sein. Grundsätzlich sollten Forschungsdaten spätestens zeitgleich mit der Veröffentlichung der entsprechenden Publikation zugänglich gemacht werden. Die Richtlinien der DA-RT Initiative (Data Access and Research Transparency), die sich im Rahmen der politikwissenschaftlichen Fachgemeinschaft APSA herausgebildet hat und zu denen sich auch die für die Kommunikationswissenschaft relevante Fachzeitschrift *Political Communication* bekennt, fordern bereits bei der Einreichung eine Zusicherung der Autor\*innen, dass die zur Publikation gehörigen Forschungsdaten, sofern sie keinen Einschränkungen unterliegen, zum Zeitpunkt der Publikation auf einem digitalen Repositorium zugänglich sind ([www.dartstatement.org](http://www.dartstatement.org)). Falls nötig, können Daten mit Embargo-Fristen versehen werden, so dass die Nachnutzung erst nach einem bestimmten, von den Datenbereitstellenden festgelegten Zeitpunkt zugänglich sind. Die Daten sollten aber dennoch zum Zeitpunkt der Veröffentlichung registriert und dieses Embargo offengelegt werden. Bei größeren, drittmittelgeförderten Projekten ist eine Veröffentlichung der Daten zeitnah nach der Erhebung anzustreben (DFG, 2009, siehe auch Schönbrodt et al., 2016). Dies scheint vor allem bei zeitsensitiven Daten wichtig, etwa zu bestimmten Ereignissen (z.B. Wahlen), um eine sinnvolle Nachnutzung zu ermöglichen.

Neben Daten, die die Grundlage einer (wissenschaftlichen) Publikation bilden, empfehlen wir explizit auch die Bereitstellung von nicht-publizierten Forschungsdaten, die gerade im Rahmen von Metaanalysen eine Rolle spielen und dem bekannten Publikations-Bias entgegenwirken können.

### 3.3 Zugang und Datenmanagement

Wie bereits in den Leitlinien der DGS (2019) ausgeführt, differenziert der RatSWD (2018) verschiedene Varianten der Archivierung von Forschungsdaten. Diese unterscheiden sich zum einen nach dem Datenzugriff, also ob eine Nutzung durch Dritte überhaupt möglich sein soll und wenn ja, durch welche Nutzer\*innengruppe, zum

anderen danach, ob die Daten nur vor Ort genutzt oder ortsunabhängig über Serverzugriff abgerufen werden können. Des Weiteren kann die Archivierung an der entsprechenden Forschungseinrichtung oder auf professionellen Repositorien erfolgen. Prinzipiell sind auch andere Formen des Archivierens und Teilens möglich, wie z.B. über das nach wie vor gängige Teilen (nur) auf Nachfrage (sharing on request) oder die Verfügbarmachung über persönliche Webseiten der Forscher\*innen. Da diese Varianten aber nicht wirklich nachhaltig sind (z.B., weil Webseiten vom Netz genommen werden oder Personen die Wissenschaft verlassen), ist von diesen Varianten generell abzuraten (Klein et al., 2018). Institutionelle oder unabhängige öffentliche Repositorien sind im Allgemeinen die deutlich bessere Lösung. Sollte es vonseiten der beteiligten Institutionen keine anderweitigen Vorgaben geben, sind unabhängige öffentliche Repositorien v.a. im Hinblick auf die Sichtbarkeit bzw. Auffindbarkeit der Daten die zu bevorzugende Variante. Klein et al. (2018) listen für die Auswahl eines geeigneten Repositoriums sieben Kriterien auf. Nach diesen sollte das Repository idealiter...

1. eindeutige und persistente Identifier für veröffentlichte Daten/Materialien verwenden (z.B. DOIs)
2. strukturierte Metadaten anbieten, um Auffindbarkeit und Nachnutzung zu vereinfachen
3. die Nutzung der Daten erfassen (z.B. über Downloads oder Zitationen)
4. die Verwendung verschiedener Lizenzen ermöglichen (siehe auch Abschnitt 4.5)
5. Zugangskontrolle erlauben
6. langfristige Archivierung garantieren
7. Daten im Einklang mit den geltenden rechtlichen Rahmenbedingungen speichern (in der EU also der Datenschutzgrundverordnung DSGVO)

Es gibt zahlreiche Repositorien mit unterschiedlichen Ausrichtungen, Schwerpunkten und Optionen. Eine umfassende und gezielte Suche nach geeigneten Repositorien ist über das Portal Re3Data möglich (<https://www.re3data.org/>). Klein et al. (2018) listen Repositorien für das Fach Psychologie auf, von denen die meisten auch für die Kommunikationswissenschaft relevant sind. Aus unserer Sicht ist es sinnvoll, kommunikationswissenschaftliche Forschungsdaten nach Möglichkeit über dieselben unabhängigen öffentlichen Repositorien zugänglich zu machen, um die Auffindbarkeit zu erhöhen und einheitliche Standards zu etablieren. In Anlehnung daran sowie an die Akademie für Soziologie (AS, 2019b) empfehlen wir zwei Archive als Standardlösungen für kommunikationswissenschaftliche Forschungsdaten, welche die oben genannten Kriterien erfüllen: Die Angebote von GESIS (<https://www.gesis.org/angebot/archivieren-und-registrieren/datenarchivierung>) für quantitative Daten und den an der Universität Bremen angesiedelten Qualiservice (<https://www.qualiservice.org/de/>) für qualitative Daten.

GESIS bietet mit SowiDataNet | datorium (<https://data.gesis.org/sharing/#!Home>) eine niedrighschwellige Lösung für das Teilen sozialwissenschaftlicher Forschungsdaten und bietet zudem verschiedene Services zur Aufbereitung und Dokumentation

für die Archivierung von Forschungsdaten angeboten an (<https://www.gesis.org/datenservices/unsere-servicepakete/daten-archivieren>). Für besonders sensible Daten bietet GESIS zudem den Zugang über ein spezielles Secure Data Center an (<https://www.gesis.org/angebot/daten-aufbereiten-und-analysieren/forschungsaufenthalte/secure-data-center-sdc>).

Auch bei Qualiservice findet sich ein umfangreiches Unterstützungsangebot, das von konkreten Leitlinien zur Erstellung eines Studienreports bis zu Informationen zur Anonymisierung von qualitativen Forschungsdaten reicht (Kretzer, 2013; vgl. weiterführend auch Kretzer & Diepenbroek, 2018; Medjedovic & Witzel, 2008).

Sowohl GESIS als auch Qualiservice bieten neben der Archivierung umfassende Beratung und Unterstützung in der Aufbereitung und Dokumentation von Forschungsdaten an. Generell empfiehlt es sich, wie bereits angeklungen, möglichst von Beginn eines Projekts an gezieltes Forschungsdatenmanagement zu betreiben – dies erleichtert auch die Übergabe der Daten an ein Repositorium. Idealerweise sollte zu diesem Zweck ein expliziter Forschungsdatenmanagementplan erstellt und befolgt werden (viele Forschungs-förderungseinrichtungen fordern dies mittlerweile auch ausdrücklich ein). Zum Forschungsdatenmanagement (für qualitative wie quantitative Daten) gibt es umfangreiche Empfehlungen und Handreichungen vom Rat für Sozial- und Wirtschaftsdaten (z.B. RatSWD, 2018) sowie Online-Materialien vom Consortium of European Social Science Data Archives (CESSDA; <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>). Einen guten Einstieg und Überblick zum Thema Forschungsdatenmanagement für Umfragedaten bietet zudem das von Jensen, Netscher und Weller (2019) herausgegebene Handbuch. Auch zur Erstellung von Forschungsdatenmanagementplänen gibt es verschiedene kostenfreie Tools für die Wissenschaft wie z.B. DMP Online (<https://dmponline.dcc.ac.uk/>) oder DataWiz (<https://datawiz.leibniz-psychology.org/DataWiz/>).

### 3.4 Förderinstrumente und Unterstützungsangebote

Die Bereitstellung von Forschungsdaten ist mit Aufwand verbunden, die es bei der Planung eines Projekts zu Bedenken gibt. Die Daten müssen durch die Forschenden so aufbereitet und dokumentiert werden, dass Nachnutzende damit arbeiten können. Trotzdem ist es möglich, dass während der Nachnutzung Kommunikation zwischen Bereitstellenden und Nachnutzenden nötig ist. Des Weiteren kann die Bereitstellung über Repositorien, vor allem wenn es sich um größere Datenmengen handelt, mit Kosten verbunden sein. Aufwand und Kosten können dabei je nach Umfang und Sensibilität des für eine Bereitstellung aufzubereitenden Datenmaterials variieren; zum Beispiel kann bei der Aufbereitung qualitativer Daten unter Umständen ein erhöhter Kommunikationsbedarf mit potentiellen Nachnutzenden bestehen, da eine umfassende Anonymisierung und das Verlorengehen des Erhebungskontextes Interpretationen erschweren und daher "Rückversicherungen" notwendig sein können.

Wie bereits aufgeführt, ist die finanzielle Unterstützung für Datenmanagement bei den meisten Drittmittelgeber\*innen mittlerweile Standard (z.B. DFG). Bei ei-

genfinanzierten Projekten muss dies in der Regel auf freiwilliger Basis geschehen. Hier appellieren wir an die jeweiligen Institute und/oder Professor\*innen, für solche Zwecke finanzielle Unterstützung einzuplanen oder zumindest die Ressourcenaufwand zu berücksichtigen und die entsprechenden die Kolleg\*innen anderweitig zu entlasten bzw. diese Tätigkeiten auch im Hinblick auf Publikation zu berücksichtigen.

Zusätzlich ist es sinnvoll, die Bereitstellung von Forschungsdaten auf unterschiedliche Arten zu incentivieren. Dies scheint schon allein deshalb wichtig, da die Generierung von Originaldaten mit ungleich mehr Aufwand verbunden ist als die Nachnutzung von Daten, was entsprechend honoriert werden sollte (siehe auch Schönbrodt et al., 2016). Auch die DFG (2015) richtet in ihren Leitlinien einen Appell an die Fachgemeinschaften, die Bereitstellung von Forschungsdaten und den dadurch entstehenden Aufwand in angemessener Weise zu würdigen. So kann durch die Nachnutzung sowohl für die Bereitstellenden als auch durch die Nachnutzenden ein Benefit entstehen.

Incentives für die Bereitstellenden erscheinen uns als wichtig, um ein Umdenken in der freiwilligen Bereitstellung von Forschungsdaten zu initiieren. Wichtig ist aber in diesem Zusammenhang auch, dass durch eine begründete Nicht-Bereitstellung von Daten, zum Beispiel wenn diese aufgrund von Datenschutzgründen oder Urheberrechtsbeschränkungen nicht weitergeben werden können, kein Nachteil entstehen darf. Entsprechend können sogenannte *Open Science Badges*, wie sie von einigen Journals bereits eingeführt wurden (z.B. *Political Communication*) zwar ein erster Anreiz sein, es muss aber klar sein, dass Publikationen ohne entsprechende Auszeichnungen nicht von mangelnder Qualität gekennzeichnet sind und dürfen entsprechend nicht als Qualitätskriterien missverstanden werden.

Ein wichtiger Aspekt im Hinblick auf die Förderung der Praxis des Datenteilens ist die Zitierbarkeit der Daten. Diese sollten den gleichen Stellenwert wie eine Publikation haben und auch entsprechend zitiert werden (vgl. dazu die Empfehlungen von GESIS: <https://www.gesis.org/angebot/daten-finden-und-abrufen/datenservice/forschungsdaten-zitieren>).

### 3.5 Rechte und Pflichten von Nachnutzenden

Um nicht nur bei der Bereitstellung, sondern auch bei der Nachnutzung von Daten maximale Transparenz zu ermöglichen, sollte die Nachnutzung an einige Bedingungen gekoppelt sein (vgl. dazu ausführlich AS, 2019b sowie Schönbrodt et al., 2016).

Grundsätzlich sind die Rechte und Pflichten der Nachnutzenden an die Art der vergebenen Lizenz geknüpft, die von Datenbereitstellenden frei und je nach Bedarf gewählt werden kann. Werden zum Beispiel Daten über eine CC-O Lizenz (<https://creativecommons.org/share-your-work/public-domain/cc0/>) bereitgestellt, könnten diese rechtlich gesehen für jegliche Zwecke und ohne Zitation der Originalpublikation oder Datenbereitstellenden genutzt werden. Der Vergabe einer entsprechenden Lizenz sollte entsprechend gut überlegt sein. In den meisten Fällen sollte eine Lizenz vergeben werden, die die unproblematische Nutzung der Daten ermöglicht, gleichzeitig aber die Erhebung der Daten durch eine entsprechende

Zitation würdigt (z.B. CC-by-Attribution 4.0 International <https://creativecommons.org/licenses/by/4.0/>). Eine ausführliche Übersicht über verschiedene Lizenzen und deren Bedeutung findet sich auch bei GESIS (<https://www.gesis.org/ssoar/home/informationen/vergabe-von-nutzungslizenzen>). Zusätzlich dazu erlauben viele Repositorien, auch die Nutzer\*innen-Gruppe der Daten zu spezifizieren (z.B. nur Wissenschaftler\*innen).

Unabhängig von der Lizenz halten wir es für empfehlenswert, dass die Datenbereitstellenden nach Möglichkeit über die Nachnutzung ihrer Daten informiert werden, insbesondere dann, wenn dies im Rahmen einer Publikation oder anderen Art von Veröffentlichung (z.B. Vortrag, Blogbeitrag, Interview) geschieht. Falls dies nicht ohnehin automatisiert über das Repositorium erfolgt, sollten die Nachnutzenden aktiv mit den Datenbereitstellenden in Kontakt treten. Darüber hinaus sollten diese über den Zweck der Nachnutzung (Reproduktion/Replikation, Metaanalyse, Re-Analyse etc.) sowie über die entsprechenden Befunde informiert werden. Dies gilt besonders, aber nicht ausschließlich, für Fälle, in denen abweichende Ergebnisse zu oder Fehler in den Originalbefunden ermittelt werden. Umgekehrt dürfen die Datenbereitstellenden in solchen Fällen einer Veröffentlichung der Befunde nicht im Wege stehen.

Die nachgenutzten Daten müssen (außer bei CC-0 Lizenzen) entsprechend zitiert werden, was optimalerweise über den im Repositorium hinterlegten Identifier passiert. Auch sollte für eine Publikation, die auf der Nachnutzung von Originaldaten besteht, kein eigener Datensatz veröffentlicht bzw. registriert werden, sondern auf die Originaldaten verwiesen werden, damit diese im Falle einer weiteren Nachnutzung im Original genutzt bzw. zitiert werden. Etwaige Veränderungen der Originaldaten (Umkodierung, Indexbildung etc.) sollte dokumentiert und berichtet werden. Wichtig ist auch, dass die Daten unter den gleichen rechtlichen und ethischen Gesichtspunkten wie die Originaldaten genutzt und veröffentlicht werden (z.B. nur Weiterverwendung zu wissenschaftlichen Zwecken). Die Datenbereitstellenden sollten diese Zwecke und ggf. Einschränkungen im Zuge der Datenveröffentlichung ebenfalls publizieren. Sollten diese nicht im Repositorium hinterlegt sein, sollten Nachnutzende die Datenbereitstellenden diesbezüglich kontaktieren.

#### **4. Fazit – Empfehlungen für die Bereitstellung von und den Umgang mit Forschungsdaten in der Kommunikationswissenschaft**

Auf Basis der bisherigen Ausführungen erscheinen uns folgende Punkte für die Bereitstellung und Nachnutzung von Forschungsdaten in der Kommunikationswissenschaft zentral:

- Die Bereitstellung und damit die Möglichkeit zur Nachnutzung von Forschungsdaten ist ein zentraler Aspekt wissenschaftlichen Erkenntnisgewinns und sollte sich als Standard im Fach etablieren. Die DGPK als Fachgesellschaft, ihre Fachzeitschriften sowie die Community sollte sich dafür umfassend einsetzen. Dazu gehört auch, den Zweck und die Möglichkeiten der Bereitstellung von Forschungsdaten sowie Strategien im Forschungsdatenmanagement in der Methodenausbildung stärker als bisher zu vermitteln und zu reflektieren.

- Die Möglichkeit zur Bereitstellung ist von verschiedenen Aspekten abhängig, etwa dem Datenschutz, der Einverständniserklärung, dem Urheberrecht sowie anderen Regelungen (z.B. in Verträgen mit Dienstleistern für die Datenerhebung). Entsprechend werden in der Kommunikationswissenschaft auch Daten generiert, die nicht für eine Nachnutzung zur Verfügbarkeit gestellt werden können.
- Die Bereitstellung von Forschungsdaten sollte unabhängig davon erfolgen, ob diese publiziert sind/werden oder ob eine Finanzierung durch öffentliche Geldgeber\*innen vorliegt.
- Wenn kommunikationswissenschaftliche Daten bereitgestellt werden, sollte dies über professionelle Repositorien geschehen, die eine sichere und komfortable Nachnutzung der Daten sowie eine Betreuung der Bereitstellenden garantieren. Empfehlenswert ist eine Einigung innerhalb der Community auf wenige Repositorien, so dass die Suche nach kommunikationswissenschaftlichen Forschungsdaten zur Nachnutzung übersichtlich ist.

Aufbauend auf diesen Kernpunkten schlägt die AG Forschungsdaten der DGPK folgende Empfehlung für verschiedene Akteursgruppen zum Umgang mit Forschungsdaten in der Kommunikationswissenschaft vor:

*Mitglieder der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft:*

Die DGPK empfiehlt ihren Mitgliedern, alle von ihnen erhobenen kommunikationswissenschaftlichen Forschungsdaten (bis auf die genannten Ausnahmen) in geeigneten Repositorien zugänglich zu machen. Dies gilt unabhängig davon, ob eine wissenschaftliche Publikation auf Basis der Daten angestrebt wird und ob diese von Dritter Seite finanziert wurden. Dies sollte zeitnah nach Abschluss der Datenerhebung, falls nötig mit einer entsprechenden Embargofrist, geschehen. Bei Publikationen sollte dies möglichst noch während des Publikations- und Reviewprozesses, spätestens aber direkt nach der Publikation, erfolgen.

*Fachzeitschriften der Deutschen Gesellschaft für Publizistik- und Kommunikationswissenschaft:*

Die DGPK empfiehlt ihren affilierten Fachzeitschriften, sich ausdrücklich für die Veröffentlichung von Forschungsdaten einzusetzen und Anreize dafür zu schaffen. Auch hier ist zu berücksichtigen, dass nicht alle Forschungsdaten einer Nachnutzung zugänglich gemacht werden können. Dies muss bei entsprechender Begründung möglich sein und solchen Studien darf im Rahmen der Begutachtung bzw. Veröffentlichung kein Nachteil entstehen.

Mögliche Maßnahmen in Anlehnung an die Transparency and Openness Promotion (TOP) Guidelines (Nosek et al., 2015 bzw. <https://cos.io/top/>) wären z.B.:

- (a) Grundlegende Verpflichtung zur Publikation der Forschungsdaten schon bei Einreichung.
- (b) "Belohnungen", z.B. in Form von Open Science Badges, um die Verfügbarkeit von Forschungsdaten prominent hervorzuheben. Hier ist allerdings zu beach-

ten, dass Studien, bei denen Daten aus nachvollziehbaren Gründen nicht verfügbar gemacht werden können, nicht abgewertet werden.

- (c) Ausdrückliche Stellungnahmen zur (Nicht-)Verfügbarkeit von Forschungsdaten bei Einreichung einzufordern und jedem Beitrag beizufügen.

Die DGPK empfiehlt ihren affilierten Fachzeitschriften, die verschiedenen Varianten der Nachnutzung von Forschungsdaten als gleichberechtigte, publikationwürdige Forschung anzusehen und somit Reproduktions-/Replikationsstudien und Sekundärdatenanalysen entsprechenden Raum zu bieten.

#### *Drittmittelgeber\*innen:*

Die DGPK empfiehlt Drittmittelgeber\*innen und Förderorganisationen, sowohl das Datenmanagement, als auch die Veröffentlichung von Forschungsdaten

- (a) explizit zu fördern, z.B. durch entsprechende Zusatzmittel,  
 (b) im Begutachtungsprozess entsprechende Praktiken auch positiv zu berücksichtigen – allerdings dürfen für Projekte, bei denen eine Veröffentlichung der Daten aus nachvollziehbaren Gründen nicht möglich ist, keine Nachteile entstehen,  
 (c) mittelfristig die Publikation der projektbezogenen Forschungsdaten, soweit möglich, zur Voraussetzung der Förderung zu machen.

## Literatur

Akademie für Soziologie (2019a). *Wissenschaftliche Daten sind kein Privateigentum einzelner Forschender, sondern ein kollektives Gut. Die Bereitstellung von Forschungsdaten zur Nachnutzung und Replikation muss auch in der Soziologie die Norm sein. Positionspapier der Akademie für Soziologie*. Abgerufen unter: <https://akademie-soziologie.de/wp-content/uploads/2019/11/AS-Stellungnahme-Nachnutzung-Forschungsdaten-01-2019.pdf> [letzter Abruf: 03.12.2020]

Akademie für Soziologie (2019b). *Richtlinien zum Umgang mit Forschungsdaten in der Soziologie*. Abgerufen unter: neuer Link: [https://akademie-soziologie.de/wp-content/uploads/2019/10/richtlinien-zum-umgang-mit-forschungsdaten-AS\\_25\\_09\\_2019-1.pdf](https://akademie-soziologie.de/wp-content/uploads/2019/10/richtlinien-zum-umgang-mit-forschungsdaten-AS_25_09_2019-1.pdf) [letzter Abruf: 03.12.2020]

Bishop, L., & Gray, D. (2017). *Chapter 7: Ethical challenges of publishing and sharing social media research data*. In K. Woodfield (Hrsg.), *Advances in Research Ethics and Integrity* (Bd. 2, S. 159–187). <https://doi.org/10.1108/S2398-60182018000002007>

DFG (2009). *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*. Abgerufen unter: [http://dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf) [letzter Abruf: 03.12.2020]

DFG (2015). *Leitlinien zum Umgang mit Forschungsdaten*. Abgerufen unter: [https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf) [letzter Abruf: 03.12.2020]

DFG (2019). Nationale Forschungsdateninfrastruktur – Ausschreibung 2019 für die Förderung von Konsortien. *Informationen für die Wissenschaft*, 37. Abgerufen unter: [https://www.dfg.de/foerderung/info\\_wissenschaft/2019/info\\_wissenschaft\\_19\\_37/index.html](https://www.dfg.de/foerderung/info_wissenschaft/2019/info_wissenschaft_19_37/index.html) [letzter Abruf: 03.12.2020]

- DGS (2019). *Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie*. Abgerufen unter: <https://soziologie.de/aktuell/news/bereitstellung-und-nachnutzung-von-forschungsdaten-in-der-soziologie> [letzter Abruf: 03.12.2020]
- Jensen, U., Netscher, S., & Weller, K. (Hrsg.). (2019). *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten*. Leverkusen: Verlag Barbara Budrich. <https://doi.org/10.3224/84742233>
- Kinder-Kurlanda, K., Weller, K., Zenk-Möltgen, W., Pfeffer, J., & Morstatter, F. (2017). Archiving information from geotagged tweets to promote reproducibility and comparability in social media research. *Big Data & Society*, Advance online publication. <https://doi.org/10.1177/2053951717736336>
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Hofelich Mohr, A., ... Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, 4, 20. <http://doi.org/10.1525/collabra.158>
- Kretzer, S. (2013). *Arbeitspapier zur Konzeptentwicklung der Anonymisierungs-/Pseudonymisierung in Qualiservice*. Abgerufen unter: <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-47605-2>
- Kretzer, S. & Diepenbroek M. (2018). Flexible Strategien für eine forschungsfreundliche Archivierung und Nachnutzung qualitativer Forschungsdaten. In RatSWD, *Archivierung und Zugang zu qualitativen Daten*. Abgerufen unter: [https://www.konsortswd.de/wp-content/uploads/RatSWD\\_WP\\_267.pdf](https://www.konsortswd.de/wp-content/uploads/RatSWD_WP_267.pdf) [letzter Abruf: 03.12.2020]
- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1, 389–402. <https://doi.org/10.1177/2515245918787489>
- Mannheimer, S., & Hull, E. A. (2018). Sharing selves: Developing an ethical framework for curating social media data. *International Journal of Digital Curation*, 12, 196–209. <https://doi.org/10.2218/ijdc.v12i2.518>
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., et al. (2016). *How open science helps researchers succeed*. *eLife*, 5, 1–19. <http://doi.org/10.7554/eLife.16800>
- Medjedovic, I., & Witzel, A. (2008). *Wiederverwendung qualitativer Daten*. Wiesbaden: VS Verlag.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348, 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Piowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, 2–25. <https://doi.org/10.7717/peerj.175>
- Schönbrodt, F. D., Gollwitzer, M., & Abele-Brehm, A. (2016). *Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien*. Abgerufen unter: [https://www.dgps.de/fileadmin/documents/Empfehlungen/Datenmanagement\\_deu\\_9.11.16.pdf](https://www.dgps.de/fileadmin/documents/Empfehlungen/Datenmanagement_deu_9.11.16.pdf) [letzter Abruf: 03.12.2020]
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531–536. <https://doi.org/10.1177/1745691612463401>
- Quintana, D. (2019). Synthetic datasets: A non-technical primer for the biobehavioral sciences [Preprint]. <https://doi.org/10.31234/osf.io/dmfb3>

- 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. Auflage). Berlin. Abgerufen unter: [https://www.ratswd.de/dl/RatSWD\\_Output3\\_Forschungsdatenmanagement.pdf](https://www.ratswd.de/dl/RatSWD_Output3_Forschungsdatenmanagement.pdf) [letzter Abruf: 03.12.2020]
- Thomson, S. D. (2016). Preserving Social Media. *DPC Technology Watch Report*. <https://doi.org/10.7207/twr16-01>
- Watteler, O., & Kinder-Kurlanda, K. E. (2015). Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung. *Datenschutz und Datensicherheit*, 39, 515–519.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3. <https://doi.org/10.1038/sdata.2016.18>
- Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, 51, 1149–1168. <https://doi.org/10.1177/0038038517708140>

## EXTENDED ABSTRACT

**Guidelines for handling research data in communication science**  
Working group research data on behalf of the management board  
of the German Communication Association (DGPUK)

*Christina Peter, Johannes Breuer, Philipp K. Masur,  
Michael Scharkow & Christian Schwarzenegger*

# EXTENDED ABSTRACT

## **Guidelines for handling research data in communication science**

Working group research data on behalf of the management board of the German Communication Association (DGPUK)

*Christina Peter, Johannes Breuer, Philipp K. Masur, Michael Scharkow & Christian Schwarzenegger*

**Acknowledgement:** In addition, we want to thank Olaf Jandura, Anna Schnauber-Stockmann, Jörg Hagenah and Corinna Lüthje for their commitment in the working group research data.

In 2016, the German Research Foundation (DFG) called upon the social science associations in Germany to address the handling of research data and to develop guidelines for their open accessibility and long-term preservation (DFG, 2015). The German Psychological Society (DGPs) was the first to develop such guidelines (Schönbrodt et al., 2016), followed by the German Sociological Association (DGS, 2019), and the Academy of Sociology (AS, 2019a, 2019b). The present guidelines for handling research data in communication science largely follows these guidelines, but takes into account the specifics of communication research and the resulting challenges for dealing with research data, especially the widespread use of content analysis as well as the prevalence of both qualitative and quantitative research in the field.<sup>1</sup>

### **The importance of research data and their accessibility for communication science**

Research data are the primary product of empirical studies and essential for generating scientific knowledge. Thus, adhering to methodological quality standards in the collection and analysis of research data is a necessary prerequisite for knowledge creation also in communication science. This applies to both qualitative and quantitative research, although the respective quality criteria may differ between these two paradigms.

Importantly, transparency is a core criterion that not only increases confidence in the results of empirical research, but, likewise, contributes to efficient cumulative knowledge creation. An important aspect in ensuring transparency is the availability and open accessibility of research data. If data are accessible to other

1 The guidelines were formulated by the Research Data Working Group, revised in several revision loops with feedback from the board of directors and division chairs, and unanimously adopted by the DGPUK members at the annual meeting of the DGPUK in Munich on March 11, 2020. The present contribution represents this final version and has accordingly not undergone a classic journal peer review.

researchers beyond the original investigators and well-documented, they can be used to investigate the robustness, reproducibility, replicability, and generalizability of the original findings. Transparency and proper documentation of the primary data also ensures the traceability of interpretative conclusions. In addition, sharing research data also increases the efficiency of research as, oftentimes, the same data can be used to answer different and novel research questions (through additional or more detailed analysis). This is not only efficient, but also ethical because a more exhaustive use of participants' data (e.g. created in surveys, observational studies, interviews, or experiments) increases the usefulness of the data and, thereby, better justifies the participants' effort. This can be particularly interesting for qualitative data whose creation often requires more time and effort both for the researchers as well as the respondents. An efficient use of research resources is further desirable as these resources are oftentimes financed by public funds. Last but not least, in the case of publicly funded projects, the results of this research should, with the necessary constraints, also be made available to the general public. In a similar vein, the Academy of Sociology refers to scientific data as a collective good (AS, 2019a, see also DFG, 2009).

From a practical perspective, systematic research data management also entails advantages for scientists themselves. Since researchers are often involved in several studies or projects and some time usually passes between the planning of a project, the execution of the study, and the publication of the results, it can be difficult for the researchers themselves to retrace or reproduce all decisions and steps in the process of collection and analysis after several weeks or months (Klein et al., 2018).

Making data available to other scientists also increases the visibility of one's own research and can open up new opportunities for collaborations and further publications (McKiernan et al., 2016). For example, a study by Piwovar and Vision (2013) showed that sharing research data leads to a higher number of citations for related publications.

## **Guidelines for sharing and reusing research data in communication science**

With the guidelines outlined here for the handling, sharing and the reuse of research data, we follow existing principles developed by the German Psychological Society (Schönbrodt et al., 2016), the German Sociological Association (DGS, 2019), and the Academy of Sociology (AS, 2019b). Furthermore, we refer to the FAIR principles for handling research data (Wilkinson et al., 2016). According to these guidelines, data should be findable, accessible, interoperable and reusable. Based on these guidelines and in addition to them, we will outline the principles we consider particularly important for dealing with research data in communication science.

## **Types of data**

We refer to research data as the so-called primary data (Schönbrodt et al., 2016), meaning the first transfer of raw data (e.g. physical questionnaires, audio recordings of interviews, journal articles) into a digital format. According to that defini-

tion, primary data are unprocessed, non-aggregated, but – as far as possible – already anonymized research data.

In contrast, providing raw data (i.e. the completed questionnaires, coded material, conversation recordings) may be difficult in many cases: These are often difficult to anonymize or extremely extensive, and complicated to reuse. Particularly in the case of qualitative interviews, raw data can often not be shared due to privacy issues.

While the protection of personal data is the primary concern in survey or observation studies, copyright issues are also relevant in many areas of communication science, especially in content analyses. Similarly, social media data often cannot be shared because the terms of service (ToS) of the platforms do not allow or at least restrict the sharing of data. This restriction applies to data obtained via Application Programming Interfaces (APIs) as well as to the purchase of data from service providers or data from direct research collaborations with the platforms (Bishop & Gray, 2017; Breuer et al., 2020; Williams et al., 2017).

In cases where primary data cannot be shared for any of these reasons, researchers should at least provide aggregated (e.g. covariance matrices) or synthetic data (Quintana, 2019). Although this does not allow direct reuse, it can be used to reproduce the results and may still provide important insights or data for meta-analyses. In general, the guiding principle for sharing research data should be “as open as possible, as restrictive as necessary,” and the effort should be based on the expected value of a given reuse of the data. The required effort results especially from the preparation or contextualization of the data sets, but also from the communication with subsequent users and should either be budgeted in third-party funded research projects (as is already provided for by the DFG, for example) or incentivized (e.g., through citations of the data via DOIs).

## Scope and timing of publication

Aspects of data protection and copyright are important aspects of sharing research data (Schönbrodt et al., 2016). First and foremost, data must be anonymized beforehand in such a way that the identification of the individual persons is not possible (at least not without disproportionate effort, Watteler & Kinder-Kurlanda, 2015). Participants’ informed consent should always be obtained, and informed consent information should include how the data will be used and shared and with whom (see recommendations by Schönbrodt et al., 2016). If it is not possible to obtain consent (e.g., when collecting social media data via Application Programming Interfaces (APIs) or purchasing such data from third-party providers), the data should be anonymized or aggregated to the extent that it is no longer possible to identify individuals.

The data provided should be prepared in such a way that it increases subsequent reusability. Therefore, published data sets should contain all variables/questions/categories that were collected in the original study, even if they were not analyzed in the publication(s) for which the data was collected (this should also be noted in the publications), since it is precisely such “unused” data that can be of interest for subsequent use. In addition, the respective measurement or instrument (i.e., questionnaire, guidelines, codebook, observation protocol, stimulus

material) should be published with the data set to ensure that subsequent users easily understand the structure and composition of the data set. If the publication of additional material is not possible, the data themselves should be labeled in a way that is comprehensible to subsequent users. If data points have been deleted from the original data set or variables and had to be removed to anonymize the data, this should be explained in a separate document describing the data set.

The timing of publishing research data may depend on the scope of the particular project as well as financing. In principle, research data should be made available with the publication of a corresponding manuscript. To that regard, the guidelines of the DA-RT Initiative (Data Access and Research Transparency) require authors to confirm at the time of submission that their data included in the publication will be accessible in a digital repository at the time of publication (see [www.dartstatement.org](http://www.dartstatement.org)). If necessary, data can be provided with embargo periods, so that subsequent use is only possible after a certain time specified by the data providers. However, the data should still be registered at the time of publication and this embargo should be disclosed. In the case of larger, third-party funded projects, the data should be shared as soon as possible after the data collection (DFG, 2009, see also Schönbrodt et al., 2016). This seems to be particularly important for time-sensitive data, (e.g. on certain events such as elections) to allow for a meaningful subsequent use.

In addition to data used in (scientific) publications, we explicitly encourage the sharing of non-published research data that play a role especially in the context of meta-analyses and can counteract potential publication bias.

## Access and data management

There are different variants of how to archive and share research data. The German Data Forum (RatSWD, 2018) differentiates these according to data access, i.e., whether it should be possible for third parties to use the data at all and, if so, by which user group(s). Researchers can decide whether the data can only be used on site (e.g. at the university of the data producer) or via remote access from any location. Furthermore, archiving can be done at the respective research institution or in professional repositories. If not specified otherwise, independent public repositories are the preferred option, especially with regard to the visibility or findability of the data. Klein et al (2018) list seven criteria for the selection of a suitable repository. According to these criteria, the repository should ideally...

1. use unique and persistent identifiers for published data/materials (e.g. DOIs)
2. offer structured metadata to facilitate retrieval and reuse
3. record the reuse of the data (e.g. via downloads or citations)
4. allow for the use of different licenses
5. allow access control
6. guarantee long-term preservation
7. store data in accordance with the applicable legal framework (e.g., the General Data Protection Regulation or GDPR in the EU)

Specifically, we recommend two repositories that fulfill these criteria: The services of GESIS (<https://www.gesis.org/en/services/archiving-and-sharing/sharing-data/data-archiving>) for quantitative data and Qualiservice (<https://www.qualiservice.org/de/>) at the University of Bremen for qualitative data.

## Incentives and support structures

The sharing of research data is associated with additional efforts that have to be considered when planning a project. The data must be prepared and documented by the researchers in a way that allows unconstrained reuse for subsequent users. Furthermore, sharing data via repositories can be associated with costs, especially if larger amounts of data are involved. The effort and costs involved can vary depending on the scope and sensitivity of the data. As mentioned above, financial support for data management has become standard for most third-party funding agencies (e.g. DFG). In the case of projects without third-party funding, this must generally be done on a voluntary basis. We appeal to the respective institutes and/or principal investigators to provide (financial) support for such purposes.

In addition, sharing research data should be incentivized in various ways. The collection of original data involves more effort than the subsequent use of data, which should be rewarded accordingly (see also Schönbrodt et al., 2016). In its guidelines, the DFG (2015) also urges the scientific communities to give appropriate recognition to the sharing of research data. Thus, the subsequent use of data can result in a benefit for both the data providers and the subsequent users.

However, it is also important that a justified refusal to share data does not result in any disadvantages. For example, privacy protection issues or copyright restrictions are valid reasons to refuse data sharing. So-called Open Science Badges, as already introduced by some journals (e.g. Political Communication), can be a first incentive, but they must not be misunderstood as quality criteria.

An important aspect in promoting the practice of data sharing is related to the citation of data. They should have the same status as publications and be cited accordingly (cf. GESIS recommendations: <https://www.gesis.org/en/services/finding-and-accessing-data/data-archive-service/citation-of-research-data>).

## Rights and obligations of subsequent users

In principle, the rights and obligations of subsequent users are determined by the type of license granted, which can be chosen by data providers in line with their needs. For example, if data is shared under a CC-O license (<https://creativecommons.org/share-your-work/public-domain/cc0/>), it could legally be used for any purpose and without citation of the original publication. The granting of such a license should be carefully considered. In most cases, a license should be granted that allows an encompassing use of the data, but at the same time acknowledges the collection of the data by citing it accordingly (e.g. CC-by-Attribution 4.0 International <https://creativecommons.org/licenses/by/4.0/>). In addition, many repositories also allow to specify who or which particular group of potential users can use the data (e.g. only scientists).

Irrespective of the license, we consider it advisable that data providers are informed about the subsequent use of their data, especially if it is used in any type of publication. The reused data must be cited accordingly (except for CC-0 licenses), ideally using persistent identifiers, such as DOIs, that many repositories provide. Also, no separate data set derived from existing published data should be published or registered. Instead, the new publication should cite the original data. Any changes to the original data (recoding, indexing, etc.) should be documented and reported. It is also important that the data is used under the same legal and ethical aspects as the original data (e.g. only reuse for scientific purposes).

### **Conclusion – Guidelines for handling and sharing research data in communication science**

Based on the discussion in this paper, the working group proposes the following recommendations for handling research data in communication science:

- The sharing and reuse of research data is a central aspect of scientific knowledge creation and should be established as a standard in the field. The DG-PuK as a professional association, its affiliated journals, and its members should make any effort to achieve this. This also includes educating students and researchers about the purpose and advantages of sharing data as well as practical research data management strategies.
- The possibility of making data available depends on various aspects, such as data protection, obtaining adequate informed consent from study participants, copyright and other regulations. Notably, communication research generates and uses data that cannot be made available for subsequent use. This must, of course, be respected.
- Research data should be made available regardless of whether it is published or financed by public funding.
- If communication research data are made available, this should be done via professional repositories that guarantee a secure and convenient reuse of the data and necessary support for the providers.

Based on these key points, the working group of proposes the following recommendations for specific actors within the DGPuK:

#### *Members of the German Communication Association (DGPuK):*

The DGPuK should recommend making all research data available in suitable repositories (with the exceptions mentioned). This applies regardless of whether a scientific publication based on the data is intended or whether it was financed by third parties. At best, the data should be shared right after the completion of the data collection. If necessary, researchers can specify an embargo period.

#### *Affiliated journals of the German Communication Association (DGPuK):*

The DGPuK should recommend its affiliated journals to explicitly support the publication of research data and to create incentives for this. Possible measures based on the Transparency and Openness Promotion (TOP) Guidelines (Nosek et

al., 2015 or <https://cos.io/top/>) would be, for example: (a) Basic obligation to publish the research data at the time of submission, (b) “Rewards”, e.g. in the form of Open Science Badges, to prominently highlight the availability of research data, (c) Requesting explicit comments on the (non-)availability of research data at the time of submission and attach them to each paper. The DGPuK should recommend its affiliated journals to consider research based on the reuse of existing data as research equally worthy of publication and, hence, to offer appropriate space for reproduction/replication studies and secondary data analysis.

### *Third-party funding:*

The DGPuK should recommend third-party funders to explicitly promote and finance both data management and publication of research data. Furthermore, such practices should be considered in the review process, and the publication of project-related research data should be made a prerequisite for funding. However, no disadvantages should arise for projects where publication of the data is not possible for justified reasons.

## References

- Academy of Sociology (AS) (2019a). *Scientific data is not the private property of individual researchers, but a collective good. The availability of research data and their subsequent use and replication must also become a norm in sociology*. Retrieved from: [https://akademie-soziologie.de/wp-content/uploads/2020/04/AS\\_Position\\_Paper\\_Reuse\\_Research\\_Data\\_1\\_2019.pdf](https://akademie-soziologie.de/wp-content/uploads/2020/04/AS_Position_Paper_Reuse_Research_Data_1_2019.pdf) [03.12.2020]
- Academy of Sociology (AS) (2019b). *Guidelines on the Provision and Handling of Research Data in Sociology*. Retrieved from: <https://osf.io/preprints/socarxiv/xbr4v/> [03.12.2020]
- Bishop, L., & Gray, D. (2017). Chapter 7: Ethical challenges of publishing and sharing social media research data. In K. Woodfield (Hrsg.), *Advances in Research Ethics and Integrity* (Bd. 2, p. 159–187). <https://doi.org/10.1108/S2398-60182018000002007>
- Breuer, J., Bishop, L., & Kinder-Kurlanda, K. (2020). The practical and ethical challenges in acquiring and sharing digital trace data: Negotiating public-private partnerships. *New Media & Society*, 22, 2058–2080. <https://doi.org/10.1177/1461444820924622>
- German Data Forum (RatSWD) (2018). *Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte* (2. Auflage) [Research data management in the social, behavioral and economic sciences. Guidance for the application and assessment of data-generating and data-using research projects]. Berlin, Germany. Retrieved from [https://www.ratswd.de/dl/RatSWD\\_Output3\\_Forschungsdatenmanagement.pdf](https://www.ratswd.de/dl/RatSWD_Output3_Forschungsdatenmanagement.pdf) [03.12.2020]
- German Research Foundation (DFG) (2009). *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten* [Recommendations for the secure storage and provision of digital primary research data]. Retrieved from [http://dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf) [03.12.2020]

- German Research Foundation (DFG) (2015). *Leitlinien zum Umgang mit Forschungsdaten* [Guidelines for handling research data]. Retrieved from [https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf) [03.12.2020]
- German Sociological Association (DGS) (2019). *Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie* [Provision and re-use of research data in sociology]. Retrieved from <https://soziologie.de/aktuell/news/bereitstellung-und-nachnutzung-von-forschungsdaten-in-der-soziologie> [03.12.2020]
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Hofelich Mohr, A., ... Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, 4, 20. <http://doi.org/10.1525/collabra.158>
- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1, 389-402. <https://doi.org/10.1177/2515245918787489>
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., et al. (2016). How open science helps researchers succeed. *eLife*, 5, 1-19. <http://doi.org/10.7554/eLife.16800>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348, 1422-1425. <https://doi.org/10.1126/science.aab2374>
- Piwovar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, 2-25. <https://doi.org/10.7717/peerj.175>
- Schönbrodt, F. D., Gollwitzer, M., & Abele-Brehm, A. (2016). *Der Umgang mit Forschungsdaten im Fach Psychologie: Konkretisierung der DFG-Leitlinien* [Dealing with research data in psychology: concretization of the DFG guidelines]. Retrieved from [https://www.dgps.de/fileadmin/documents/Empfehlungen/Datenmanagement\\_deu\\_9.11.16.pdf](https://www.dgps.de/fileadmin/documents/Empfehlungen/Datenmanagement_deu_9.11.16.pdf) [03.12.2020]
- Quintana, D. (2019). Synthetic datasets: A non-technical primer for the biobehavioral sciences [Preprint]. <https://doi.org/10.31234/osf.io/dmfb3>
- Watteler, O., & Kinder-Kurlanda, K. E. (2015). Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung [Anonymization and secure handling of research data in empirical social research]. *Datenschutz und Datensicherheit*, 39, 515-519.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3. <https://doi.org/10.1038/sdata.2016.18>
- Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, 51, 1149-1168. <https://doi.org/10.1177/0038038517708140>