

Sozialwissenschaftliche Forschungsdaten langfristig sichern und zugänglich machen: Herausforderungen und Lösungsansätze

Anja Perry, *GESIS – Leibniz-Institut für Sozialwissenschaften, Köln*

Jonas Recker, *GESIS – Leibniz-Institut für Sozialwissenschaften, Köln*

Zusammenfassung:

Sozialwissenschaftlerinnen und -wissenschaftler sind gemäß der guten wissenschaftlichen Praxis dazu angehalten, den Forschungsprozess möglichst transparent zu halten, ihre Forschungsergebnisse reproduzierbar zu machen und im Sinne von „Open Science“ erhobene Daten für die Nachnutzung zur Verfügung zu stellen. Dies ist allerdings mit einigen Hürden verbunden. So müssen nachnutzbare Daten so dokumentiert werden, dass sie für Dritte auffindbar und verständlich sind. Gleichzeitig müssen personenbezogene Daten der Teilnehmenden ausreichend geschützt werden. Forschende sind hier auf Unterstützung angewiesen, um diesen Ansprüchen an die empirische Forschung gerecht zu werden. Dieser Artikel geht zunächst auf die Hürden bei der Bereitstellung von sozialwissenschaftlichen Forschungsdaten ein und stellt dann Dienstleistungen des GESIS Datenarchivs für Sozialwissenschaften vor, die Forschenden helfen, diese Hürden zu meistern.

Summary:

In accordance with the rules of good scientific practice, social scientists are encouraged to keep the research process as transparent as possible, to make their research results replicable, and – in the spirit of “Open Science” – to make collected data available for re-use. However, this is associated with a number of hurdles. For example, reusable data must be documented in such a way that it can be found and understood by third parties. At the same time, personal data of respondents must be adequately protected. Researchers need support to meet these demands on empirical research. The article first discusses the barriers to the provision of social science research data and afterwards presents special services of the GESIS Data Archive for Social Sciences, which help researchers to overcome these hurdles.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2018H2S106-122>

Autorenidentifikation: Anja Perry: ORCID <https://orcid.org/0000-0003-0574-9275>; Jonas

Recker: ORCID <https://orcid.org/0000-0001-9562-3339>

Schlagwörter: Sozialwissenschaften, Forschungsdaten, Datenmanagement, Archivierung

1. Einleitung

Die Anforderungen an Wissenschaftlerinnen und Wissenschaftler, möglichst umfassend Transparenz über den Forschungsprozess und die erzielten Ergebnisse herzustellen und damit Forschungsergebnisse reproduzierbar und nachvollziehbar zu machen, sind in den letzten Jahren stetig gestiegen. Enthielten schon die Regeln zur Sicherung der guten wissenschaftlichen Praxis der Deutschen

Forschungsgemeinschaft¹ Forderungen mit Blick auf eine Aufbewahrung von „Primärdaten“ inklusive einer Beschreibung der Erhebungsmethoden, so knüpfen Drittmittelgeber Förderzusagen heute zunehmend an das Vorliegen eines geeigneten Datenmanagementplans, welcher im Projektverlauf umgesetzt wird und idealerweise in einer Bereitstellung der Forschungsdaten für die Nachnutzung mündet. Das BMBF fordert in einzelnen Programmen beispielsweise, dass „Standards des Forschungsdatenmanagements“ eingehalten und erhobene Forschungsdaten nach Projektende an ein geeignetes Forschungsdatenzentrum übergeben werden, „um im Sinne der guten wissenschaftlichen Praxis eine langfristige Datensicherung für Replikationen und gegebenenfalls Sekundärauswertungen zu ermöglichen“.² Vergleichbare Forderungen oder mindestens Empfehlungen finden sich bei der DFG und der EU (Horizon 2020).³

Auch wissenschaftliche Zeitschriften beziehen die einer Publikation zugrundeliegenden Forschungsdaten immer häufiger in den Peer-Review-Prozess mit ein und empfehlen oder fordern die Veröffentlichung dieser Daten in geeigneten Systemen.⁴

Im Zuge dieser Entwicklung ist ein Bedarf an Forschungsdateninfrastrukturen entstanden, d.h. an Einrichtungen, die (fachspezifische) „Instrumente, Ressourcen und Dienstleistungen zur Verfügung [stellen]“.⁵ Dieser Bedarf wird mittlerweile auf nationaler und internationaler Ebene von einer Vielzahl an Initiativen und Projekten adressiert, in deren Rahmen disziplinspezifische und generische Angebote zum Management, der Publikation und der langfristigen Archivierung von Forschungsdaten entstehen. Hochschulen spielen hierbei neben disziplinspezifischen Infrastrukturanbietern eine wichtige Rolle: Sie vermitteln Forschenden Wissen und Kompetenzen für den Umgang mit Forschungsdaten; sie erarbeiten institutionelle Forschungsdatenpolicies und -leitlinien und setzen diese um; sie etablieren Beratungs- und Unterstützungsangebote für Forschende und stellen Systeme zur Veröffentlichung von Forschungsdaten bereit, häufig in enger Zusammenarbeit zwischen Bibliotheken, Rechenzentren und der Hochschulleitung. Beispiele hochschulischer Initiativen auf Landesebene sind die Projekte „bwFDM – Koordiniertes Forschungsdatenmanagement in Baden-Württemberg“ (abgeschlossen)

1 Vgl. Deutsche Forschungsgemeinschaft, *Sicherung guter wissenschaftlicher Praxis / Safeguarding Good Scientific Practice*. Denkschrift / Memorandum, 2013, Empfehlung 7, zuletzt geprüft am 09.03.2018, http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.

2 Bundesministerium für Bildung und Forschung, „Bekanntmachung Richtlinie zur Förderung von Forschung zu ‚Digitalisierung im Bildungsbereich – Grundsatzfragen und Gelingensbedingungen‘“, Bundesanzeiger, Nr. AT 26.09.2017 B2 (2017), zuletzt geprüft am 14.02.2018, <https://www.bmbf.de/foerderungen/bekanntmachung-1420.html>.

3 Deutsche Forschungsgemeinschaft, *Umgang mit Forschungsdaten: DFG-Leitlinien zum Umgang mit Forschungsdaten*, 2015, zuletzt geprüft am 27.06.2018, http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/;

European Commission, *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*, 2016, zuletzt geprüft am 14.02.2018, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

4 Beispiele sind PLOS (<http://journals.plos.org/plosone/s/data-availability>) und die Zeitschrift für Soziologie (<https://www.degruyter.com/view/j/zfsoz#callForPapersHeader>). Links zuletzt geprüft am 14.02.2018.

5 Wissenschaftsrat, *Bericht zur wissenschaftsgeleiteten Bewertung umfangreicher Forschungsinfrastruktur-vorhaben für die Nationale Roadmap*, 2017, 8, zuletzt geprüft am 14.02.2018, <https://www.wissenschaftsrat.de/download/archiv/6410-17.pdf>.

und „Hessische Forschungsdateninfrastrukturen – HeFDI“; auch in Nordrhein-Westfalen besteht seit 2016 eine einrichtungsübergreifende Kooperation.⁶

Eine Herausforderung, vor der solche hochschulischen Angebote stehen, ist, dass sie eine große Bandbreite an wissenschaftlichen Fachrichtungen mit ganz unterschiedlichen Voraussetzungen und Anforderungen bezüglich Beratung oder technischer Infrastruktur bedienen müssen. In der empirischen Sozialforschung spielen hier insbesondere datenschutzrechtliche Fragen eine Rolle, da häufig so genannte personenbezogene Daten erhoben und verarbeitet werden. Auch der Erhalt der Interpretierbarkeit der Daten, der maßgeblich von der Verfügbarkeit ausreichender Kontextinformationen zur Datenerhebung und -verarbeitung abhängt, folgt disziplinspezifischen Praktiken und Standards.

Oft ist es nicht möglich, im Rahmen einer generischen Infrastruktur allen fachspezifischen Anforderungen umfassend und in ausreichender Tiefe gerecht zu werden. Hierbei können disziplinspezifische Infrastrukturen wie das GESIS – Leibniz-Institut für Sozialwissenschaften wichtige Aufgaben übernehmen, indem sie fachbezogene Unterstützungs- und Lösungsangebote zur langfristigen Sicherung und Bereitstellung von Forschungsdaten bieten und damit generische Angebote ergänzen. Nach einer kurzen Beschreibung, was unter sozialwissenschaftlichen Forschungsdaten verstanden wird, soll im folgenden Beitrag vor diesem Hintergrund ein Bewusstsein für Herausforderungen geschaffen werden, die bei der langfristigen Sicherung und Bereitstellung sozialwissenschaftlicher Forschungsdaten zu beachten sind. Anschließend werden beispielhaft Unterstützungs- und Lösungsangebote von GESIS – Leibniz-Institut für Sozialwissenschaften skizziert.

2. Was sind sozialwissenschaftliche Daten?

Forschende in den Sozialwissenschaften arbeiten mit einer großen Bandbreite unterschiedlicher Arten von Daten. Eine typische Unterscheidung wird dabei zwischen quantitativen und qualitativen Daten getroffen. Quantitative sozialwissenschaftliche Daten sind häufig Umfragedaten. Das können Meinungsumfragen sein, wie das Politbarometer,⁷ Vergleichsstudien, wie das Programme for the International Comparison of Adult Competencies (PIAAC) zum internationalen Vergleich der Kompetenzen Erwachsener,⁸ oder auch Langzeitbefragungen, wie das Beziehungs- und Familienpanel pairfam.⁹ Auch Prozess- oder Registerdaten, wie z.B. die Verwaltungsdaten der sozialversicherungspflichtig Beschäftigten bei der Bundesagentur für Arbeit, werden für Auswertungen verwendet. Aus dieser Bandbreite an Daten ziehen Sozialwissenschaftlerinnen und Sozialwissenschaftler Rückschlüsse über die Gesellschaft und geben Handlungsempfehlungen an die Politik.

6 „bwFDM - Koordiniertes Forschungsdatenmanagement in Baden-Württemberg,“ zuletzt geprüft am 14.02.2018, <https://bwfdm.scc.kit.edu/>; „Hessische Forschungsdateninfrastrukturen - HeFDI,“ zuletzt geprüft am 14.02.2018, <https://www.uni-marburg.de/projekte/forschungsdaten/projekt>; Constanze Curdt et al., „Herausforderung Forschungsdatenmanagement – Unterstützung der Hochschulen durch eine einrichtungsübergreifende Kooperation in NRW,“ in *E-Science-Tage 2017: Forschungsdaten managen*, hrsg. Jonas Kratzke und Vincent Heuveline (Heidelberg: heiBOOKS, 2017), <https://doi.org/10.11588/heibooks.285.377>.

7 <https://doi.org/10.4232/1.12824>.

8 <http://www.oecd.org/skills/piaac/>.

9 <https://doi.org/10.4232/pairfam.5678.8.0.0>.

Die Daten sind in der Regel in einer Datenmatrix organisiert und stark strukturiert. In dieser Datenmatrix sind die befragten Personen in Zeilen angeordnet (z.B. eine Zeile pro befragter Person) und die einzelnen Variablen spaltenweise angeordnet. Oftmals liegen diese Daten in proprietären Formaten, wie den gängigen Statistikformaten SPSS und Stata, oder aber auch als Tabelle (MS Excel) oder im Textformat (csv) vor (s. Abbildung 1).

	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
1	20001 1.0.0 (2017-07-01)		1	2	5	2	1	2	99	1	1
2	20002 1.0.0 (2017-07-01)		1	2	5	2	5	2	2	1	1
3	20003 1.0.0 (2017-07-01)		1	2	17	1	8	1	11	1	1
4	20004 1.0.0 (2017-07-01)		1	2	5	2	4	2	29	1	2
5	20005 1.0.0 (2017-07-01)		1	2	15	2	5	3	35	1	1
6	20006 1.0.0 (2017-07-01)		1	2	8	2	4	46	99	1	5
7	20007 1.0.0 (2017-07-01)		1	2	16	1	1	2	7	1	2
8	20008 1.0.0 (2017-07-01)		1	2	15	1	5	2	24	1	6
9	20009 1.0.0 (2017-07-01)		1	2	17	1	3	2	99	2	0
10	20010 1.0.0 (2017-07-01)		1	2	14	1	4	2	29	1	2
11	20011 1.0.0 (2017-07-01)		1	2	5	2	8	2	8	1	99
12	20012 1.0.0 (2017-07-01)		1	2	17	1	1	2	99	1	3
13	20013 1.0.0 (2017-07-01)		1	2	6	2	2	38	2	1	99
14	20014 1.0.0 (2017-07-01)		1	2	15	1	6	2	35	2	99
15	20015 1.0.0 (2017-07-01)		1	2	2	2	8	2	14	1	2
16	20016 1.0.0 (2017-07-01)		1	2	12	1	8	2	36	1	3
17	20017 1.0.0 (2017-07-01)		1	2	5	2	7	2	32	1	2
18	20018 1.0.0 (2017-07-01)		1	2	15	1	2	2	3	1	99
19	20019 1.0.0 (2017-07-01)		1	2	16	1	7	2	40	2	0
20	20020 1.0.0 (2017-07-01)		1	2	1	2	1	2	99	1	1

Abbildung 1: Beispiel einer Datenmatrix in SPSS

Qualitative Daten hingegen sind nicht-numerische Daten, die häufig schriftlich (z.B. Interviewtranskripte, Beobachtungsprotokolle, Feldtagebücher oder Texte in Tageszeitungen) oder als Audio- oder Videoaufzeichnungen (z.B. von Gruppendiskussionen oder Beobachtungsstudien) vorliegen. Diese sind, im Gegensatz zu quantitativen Daten, daher oft nicht stark strukturiert.

3. Der Übergang sozialwissenschaftlicher Daten in die Langzeitsicherung: Herausforderungen und Lösungen

3.1. Die Interpretierbarkeit sicherstellen

Daten die anderen Forschenden zur Nutzung bereitgestellt werden, müssen für diese langfristig interpretierbar sein. Dies erfordert eine gründliche Aufbereitung und Dokumentation der Daten, da diese nicht für sich selbst sprechen. Es sind also zusätzliche Informationen nötig, die die Daten beschreiben und damit interpretierbar machen. Diese werden den Daten in Form von strukturierten oder unstrukturierten Metadaten („Dokumentation“) zur Seite gestellt und dokumentieren beispielsweise, welche Personen unter welchen Bedingungen befragt wurden.

Im Fall von quantitativen Daten unterscheidet man in Dokumentation auf Studienebene und auf Variablenebene. Die Dokumentation auf Studienebene umfasst, wo zutreffend, den Fragebogen, einen Methodenbericht sowie weitere Dokumente, die zum Verstehen der Studie als Ganzes wichtig sein könnten. Hier können Besonderheiten bei der Erhebung eine Rolle spielen. Die Einstellungen zur Atomkraft können sich beispielsweise kurz nach dem Reaktorunfall von Fukushima stark verändert haben und in Befragungen zu diesem Thema zu starken Trendänderungen führen. Der Fragebogen

erlaubt, die Originalfrage zu betrachten und Rückschlüsse auf das Antwortverhalten zu ziehen. Der Methodenbericht gibt einen Überblick über Erhebungsdesign, Erhebungsprozess und Datenverarbeitung.¹⁰ Zum Beispiel ist die Antwortrate ein entscheidendes Merkmal, um einschätzen zu können, wie repräsentativ die Erhebung ist. Beide Dokumente erlauben somit Dritten, die Qualität der Erhebung und ggf. die Nutzbarkeit der Daten für ihre eigenen Forschungsinteressen zu bewerten.

Auf Variablenebene fokussiert die Dokumentation auf die Beschreibung, die Struktur und den Inhalt der Daten. Ziel ist es, den Nutzenden einen schnellen Überblick über die Daten zu geben. Zentral sind hierbei die Dokumentation der Variablen im Datensatz selbst (durch Variablenname und -label) sowie das Codebook. Darin sind alle Variablen mit Namen, Label, Format, Ausprägungen (inkl. der Kennzeichnung fehlender Werte) und zugrundeliegenden Konzepten aufgelistet. Es stellt damit die Verbindung zwischen Variablen im Datensatz und Originalvariable im Fragebogen her.¹¹

Im Fall von qualitativen Daten müssen ebenfalls auf Studienebene Kontextinformationen, wie Ort der Erhebung und Zeitgeschehnisse, dokumentiert werden. Außerdem sollten beispielsweise im Fall von Interviews der Interviewleitfaden, ursprüngliche Fragenformulierungen, Anweisungen an die interviewenden Personen sowie zusätzliche Erklärungen nachgehalten werden. Im Fall von Ethnographie und (teilnehmenden) Beobachtungsstudien sind dies z.B. Feldtagebücher oder Auszüge daraus. Auf Fallebene muss erkennbar sein, wie verschiedene Dateien (z.B. Audiomitschnitt und Transkription) zueinander stehen, wer die Interviews oder Videoaufnahmen transkribiert hat und welche Situationsmerkmale in den jeweiligen Interviews eine Rolle gespielt haben.¹²

Um eine gute Aufbereitung und Dokumentation sicherzustellen, wird empfohlen, von Beginn des Forschungsprojekts an einen Datenmanagementplan zu verfassen.¹³ Dieser legt Verantwortliche für bestimmte Aufgaben im Projekt fest und definiert Abläufe für diese Aufgaben. Er wird im Laufe des Projekts immer wieder angepasst, um neu entstehenden Problemen und Aufgaben im Projekt gerecht zu werden.

10 Vgl. Oliver Watteler, *Erstellung von Methodenberichten für die Archivierung von Forschungsdaten*, 2010, zuletzt geprüft am 14.02.2018, https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Aufbau_Methodenbericht_v1_2010-07.pdf; Verbund Forschungsdaten Bildung, *Hinweise zur Erstellung eines Methoden- / Feldberichts*, fdbinfo Nr. 3, 2015, zuletzt geprüft am 14.02.2018, https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdbinfo_3.pdf.

11 Vgl. z.B. Uwe Jensen, „Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten,“ *GESIS Technical Reports*, Nr. 7 (2012), <http://nbn-resolving.de/urn:nbn:de:O168-ssoar-320650>; Karoline Harzenetter, „Metadata documentation at the variable level,“ in *Data Processing and Documentation: Generating High Quality Research Data in Quantitative Social Science Research* (GESIS Papers), hrsg. Sebastian Netscher und Christina Eder (Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, in Vorbereitung).

12 Vgl. Cabrera und Haake, „Das Qualiservice Metadatenschema. Version 1.0,“ *Qualiservice Technical Reports*, Nr. 13 (2013), <http://nbn-resolving.de/urn:nbn:de:gbv:46-00103310-19>.

13 Vgl. z.B. Ulf Jakobsson, Ricarda Braukmann und Malin Lundgren, „Expert tour guide on Data Management. Chapter 1: Plan,“ 2017, zuletzt geprüft am 14.02.2018, <http://www.cessda.eu/DMGuide>; Rat für Sozial- und Wirtschaftsdaten, *Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte*, RatSWD Output Series 3 (Berlin, 2016), <https://doi.org/10.17620/02671.7>.

3.2. Ethisch-rechtliche Anforderungen

Die Erhebung, Aufbereitung und Archivierung von Umfrage- und Registerdaten sowie qualitativen Daten betrifft immer auch personenbezogene Daten. Personenbezogene Daten sind „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person [...] beziehen“.¹⁴ Diese zu schützen, um Teilnehmende vor negativen Folgen der Forschung sowie einer kommerziellen Nutzung der Daten zu bewahren, ist ein Grundrecht der Teilnehmenden nach der Europäischen Menschenrechtskonvention.¹⁵ Aus diesem Grund müssen sich Sozialwissenschaftlerinnen und Sozialwissenschaftler möglichst frühzeitig auch mit dem Datenschutz auseinandersetzen. Dies sollte schon bei der Fragebogenerstellung geschehen, indem im Sinne der Datensparsamkeit nur Fragen gestellt werden, die tatsächlich für die Forschungsfrage relevant sind (Datenminimierung).¹⁶

Zudem muss für die Erhebung personenbezogener Daten eine Einverständniserklärung eingeholt werden.¹⁷ In Tabelle 1 werden die wichtigsten Bestandteile der informierten Einverständniserklärung zusammengefasst. In aller Regel wird das informierte Einverständnis als freiwillige aktive Zustimmung zur Teilnahme vor Beginn der Datenerhebung eingeholt. In Ausnahmefällen, z.B. wenn der Zweck der Forschung den Teilnehmenden nicht vorab bekannt sein darf, kann diese auch im Nachhinein eingeholt werden.¹⁸ Der Informationsteil mit den in Tabelle 1 genannten Bestandteilen sollte bei den Teilnehmenden verbleiben. Das deutsche Recht verlangte bisher eine schriftliche Einverständniserklärung. Laut der neuen europäischen Datenschutzgrundverordnung (DSGVO) ist eine schriftliche Einverständniserklärung seit dem 25.05.2018 nicht mehr zwingend nötig. Jedoch müssen Forschende nachweisen können, dass sie die informierte Einwilligung eingeholt haben.¹⁹ Dafür ist weiterhin eine schriftliche Einwilligung am besten geeignet. Bei bestimmten Gruppen von Teilnehmenden kann und muss man allerdings davon abweichen. Zum Beispiel können Personen ohne ausreichende Lese-/Schreibkompetenz die informierte Einwilligung nur mündlich geben.

14 Art. 4 „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 (DSGV)“ (2016), zuletzt geprüft am 14.02.2018, <http://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679&from=DE>.

15 Vgl. Art. 8 Abs. 1 „Europäische Menschenrechtskonvention (Konvention zum Schutze der Menschenrechte und Grundfreiheiten) vom 04.11.1950,“ zuletzt geprüft am 14.02.2018, <https://dejure.org/gesetze/MRK>.

16 Vgl. Art. 5 Abs. 1 DSGVO.

17 Vgl. Art. 13 Abs. 1 ebd.

18 Vgl. Deutsche Gesellschaft für Soziologie und Berufsverband Deutscher Soziologinnen und Soziologen, „Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS),“ 2017, §2 Abs. 3, zuletzt geprüft am 10.03.2018, <http://www.soziologie.de/de/die-dgs/ethik/ethik-kodex.html>.

19 Vgl. Art. 7 Abs. 1 ebd.; für eine Zusammenfassung der für die Forschung relevante Neuerungen durch die DSGVO vgl. Katrin Schaar, *Was hat die Wissenschaft beim Datenschutz künftig zu beachten? Allgemeine und spezifische Änderungen beim Datenschutz im Wissenschaftsbereich durch die neue Europäische Datenschutzgrundverordnung*, Working Paper Series des Rates für Sozial- und Wirtschaftsdaten, Nr. 257 (2016), <https://doi.org/10.17620/02671.19>.

Tabelle 1: Bestandteile einer informierten Einverständniserklärung

Inhaltlicher Informationsteil	<ul style="list-style-type: none">• Projekttitel• Zweck der Studie• Namen der beteiligten Forschenden• Erhebung personenbezogener Daten• Datennutzung im Projekt und über das Projekt hinaus• Aufbewahrungsfristen für die Daten• Einsichtsrechte und Weitergabe an Dritte (z.B. Archivierung und Nachnutzung)• Schutz persönlicher Angaben (z.B. durch Pseudonymisierung)• „in präziser, transparenter, verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache“²⁰
Rechtlicher Informationsteil	<ul style="list-style-type: none">• Freiwilligkeit der Teilnahme• Informationsrecht• Recht auf Rücknahme der Einwilligung zu jeder Zeit• Recht auf Einsichtnahme in die persönlichen Daten• Recht auf Verbesserung der persönlichen Daten• Recht auf Löschung der persönlichen Daten
Einwilligungs- erklärung	<ul style="list-style-type: none">• Besonderheiten der Befragung (z.B. Videoaufzeichnung) müssen abgedeckt sein• Stabile Kontaktinformation für Rückfragen und Wahrnehmung der zugesicherten Rechte• Unterschrift

Auch für die Archivierung und Nachnutzung der erhobenen Daten ist die informierte Einwilligung wichtig. Fehlt diese, fehlt die rechtliche Grundlage für die Datenerhebung und damit auch für die Aufbewahrung und Nachnutzung der Daten. Aber auch eine unglückliche Formulierung, wie z.B. die Zusicherung, dass die Daten nur im Forschungsprojekt verwendet werden, schließen eine Archivierung und eine Nutzung über den Zweck des Forschungsprojektes hinaus aus. Die neue DSGVO verlangt präzise und transparente Informationen über die Erhebung, Weiterverarbeitung und Aufbewahrung der Daten.²¹

Forschungsdaten müssen für die Weiterverarbeitung anonymisiert werden. Hier galt bislang in Deutschland der Grundsatz der faktischen Anonymisierung.²² Dabei dürfen die Einzelangaben „nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer

20 Art. 12 Abs. 1 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 (DSGV).

21 Vgl. Schaar, *Was hat die Wissenschaft beim Datenschutz künftig zu beachten?*

22 Vgl. Heike Wirth, „Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes.“ ZUMA Nachrichten 16, Nr. 30 (1992): 7-65, <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-209679>.

bestimmten oder bestimmbar natürlichen Person zugeordnet werden können“.²³ Nach einer Stellungnahme der Artikel-29-Datenschutzgruppe greift die neue DSGVO die faktische Anonymisierung auf und verlangt bei der Anonymisierung oder Pseudonymisierung eine Abwägung, ob eine Person bestimmbar ist. Bei dieser Abwägung müssen Kosten der Identifizierung, erforderlicher Zeitaufwand sowie verfügbare Technologien und technologische Entwicklungen berücksichtigt werden.²⁴

Bei der Pseudonymisierung wird in direkte und indirekte Identifikatoren unterschieden. Wie diese definiert sind und wie mit ihnen umgegangen wird, um die erhobenen Daten zu anonymisieren, ist in Tabelle 2 dargestellt. Alle Maßnahmen zur Pseudonymisierung müssen sorgfältig für die Datennutzer/innen dokumentiert werden, um falsche Rückschlüsse aus den Daten zu vermeiden.

Tabelle 2: Direkte und indirekte Identifikatoren und Pseudonymisierungsstrategien²⁵

Art des Identifikators	Definition	Pseudonymisierungsstrategie
Direkt	Angaben, die Teilnehmende direkt identifizieren, wie z.B. Name und Kontaktdaten, Bild oder Stimme bei audiovisuellen Medien	<ul style="list-style-type: none"> • Schnellstmögliche, dauerhafte Löschung dieser Angaben nach der Befragung • Verpixelung und/oder Verzerrung von Bild-/Tondokumenten • Verwendung von Substituten, wie z.B. „Schüler 1“ • Erfolgt eine Wiederbefragung, müssen sie getrennt von den erhobenen Daten sicher und unzugänglich für Dritte aufbewahrt werden

23 § 3 Abs. 6 „Bundesdatenschutzgesetz (BDSG)“, zuletzt geprüft am 27.06.2018, https://dejure.org/gesetze/BDSG_a.F./3.html.

24 Vgl. Erwägungsgrund Nr. 26 DSGVO; Article 29 Data Protection Working Party, „Opinion 05/2014 on Anonymisation Techniques“, 2014, zuletzt geprüft am 16.02.2018, http://collections.internetmemory.org/haeu/20140424090503/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.

25 Vgl. Heike Wirth, *Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific-Use-File*, 2006, zuletzt geprüft am 14.02.2018, <https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier11.html>; Oliver Watteler und Katharina E. Kinder-Kurlanda, „Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung,“ *Datenschutz und Datensicherheit : DuD ; Recht und Sicherheit in Informationsverarbeitung und Kommunikation* 39, Nr. 8 (2015): 515–19, <https://doi.org/10.1007/s11623-015-0462-0>; Alexia Meyermann und Maïke Porzelt, *Hinweise zur Anonymisierung von qualitativen Daten*, forschungsdaten bildung informiert, Nr. 1, 2014, zuletzt geprüft am 14.02.2018, https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-1.pdf; Marcus Eisentraut, „Data anonymization,“ in *Data Processing and Documentation: Generating High Quality Research Data in Quantitative Social Science Research* (GESIS Papers), hrsg. von Sebastian Netscher und Christina Eder (Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, in Vorbereitung).

Indirekt	Informationen, durch die in Kombination miteinander Personen identifiziert werden können, z.B. weil sie sehr kleinteilig sind (wie der Wohnort) oder Merkmale, die nur sehr selten in der Grundgesamtheit auftreten (wie ein bestimmtes Herkunftsland)	<ul style="list-style-type: none">• Vergrößerung regionaler Informationen• Zusammenfassen von Merkmalen in Kategorien• Aggregieren von Informationen
-----------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Nicht zuletzt müssen personenbezogene Daten auch vor unautorisiertem Zugriff durch Dritte geschützt werden.²⁶ Dies kann durch Zugangskontrolle über ein sicheres Passwort oder durch Verschlüsselung der Daten erfolgen. Daten die nicht mehr verwendet werden, sollten schnellstmöglich dauerhaft gelöscht werden.

4. Sicherung und Bereitstellung von Umfragedaten am GESIS Datenarchiv für Sozialwissenschaften

Gegründet im Jahr 1960 als Zentralarchiv für empirische Sozialforschung ist das Datenarchiv für Sozialwissenschaften heute eine Abteilung von GESIS – Leibniz-Institut für Sozialwissenschaften, der größten europäischen Infrastruktureinrichtung für die Sozialwissenschaften (www.gesis.org). GESIS unterstützt Wissenschaftlerinnen und Wissenschaftler, die mit Methoden der quantitativen Sozialforschung arbeiten, in allen Phasen des Forschungsprozesses. Zu den forschungsbasierten Angeboten und Dienstleistungen gehört neben der methodischen Beratung zur Studienplanung und -implementierung, der Durchführung kognitiver Pretests oder der Ziehung von Telefonstichproben zum Beispiel auch die Unterstützung von Forschenden und Forschungseinrichtungen beim Management und der langfristigen Sicherung sowie der Bereitstellung quantitativer sozialwissenschaftlicher Forschungsdaten.²⁷ So stehen im GESIS Datenarchiv über 5500 Studien für die Nachnutzung zur Verfügung und können in der Regel nach einer Registrierung im Datenbestandskatalog DBK direkt für die eigene Forschung und Lehre heruntergeladen werden.²⁸ Forschungsdatenzentren wie z.B. das FDZ ALLBUS oder das FDZ Internationale Studien sorgen für eine added-value Aufbereitung und Bereitstellung von Daten aus großen Studienprogrammen.²⁹

4.1. Übernahme von quantitativen Forschungsdaten in das GESIS Datenarchiv für Sozialwissenschaften

Um den oben skizzierten Herausforderungen bei der langfristigen Sicherung und Bereitstellung von sozialwissenschaftlichen Forschungsdaten für die Nutzung zu begegnen, führt das GESIS Datenarchiv für Sozialwissenschaften bei der Übernahme solcher Daten eine Reihe von Überprüfungen durch. Diese sind geleitet von folgenden zentralen Fragestellungen:

²⁶ Vgl. Art. 5 Abs. 1 DSGVO.

²⁷ Weitere Informationen zum Angebot von GESIS sind unter <https://www.gesis.org/angebot/> abrufbar.

²⁸ Datenbestandskatalog, <https://dbk.gesis.org/dbksearch?db=d>, zuletzt geprüft am 14.02.2018.

²⁹ GESIS Forschungsdatenzentren, <https://www.gesis.org/institut/forschungsdatenzentren/>, zuletzt geprüft am 14.02.2018.

- Liegen ausreichend Informationen vor, sodass Dritte die eingereichten Forschungsdaten verstehen und nutzen können?
- Sind die Forschungsdaten von guter Qualität (d.h. z.B. fehlerfrei aufbereitet)?
- Gibt es technische Probleme bei der Nutzung der eingereichten Forschungsdaten und Dokumentation?
- Bestehen rechtliche Probleme im Hinblick auf den Datenschutz oder das Urheberrecht?

Um diese Fragen zu beantworten, führen erfahrene Kuratorinnen und Kuratoren umfassende Überprüfungen durch (siehe Tabelle 3).

Tabelle 3: Überprüfungen bei Dateneingang im GESIS Datenarchiv (Auswahl)

technisch	<ul style="list-style-type: none">• Lesbarkeit und Virenfreiheit aller Dateien
studienbezogen	<ul style="list-style-type: none">• Vollständigkeit der eingereichten Dokumente (maschinenlesbarer Datensatz, Codebuch, Erhebungsinstrument, Methodenbericht);• die Einhaltung des Datenschutzes (sind Daten ausreichend anonymisiert?) sowie des Urheberrechts (wurden z.B. urheberrechtlich geschützte Instrumente verwendet?).
datensatzbezogen	<ul style="list-style-type: none">• Vollständigkeit der Fälle• Vollständigkeit und Korrektheit der Variablen, Variablennamen und Variablenlabel (auch: passen die Variablen zum Fragebogen?)• Gewichte• Plausibilität (u.a. Verteilung nach Geschlecht und Alter)• undokumentierte Werte (so genannte „wild codes“), fehlende Werte

Sofern hier Probleme deutlich werden, wird in Absprache mit den datengebenden Forscherinnen und Forschern versucht, diese auszuräumen – z.B. indem die Pseudonymisierung angepasst wird. Sollten Inkonsistenzen auffallen, die in einer fehlerhaften Verarbeitung der Daten begründet sind, erfolgt in Rücksprache mit den datengebenden Personen eine Korrektur. Inkonsistenzen die sich nicht korrigieren lassen – weil sie z.B. in einem verzerrten Antwortverhalten begründet sind – werden dokumentiert, damit Forschende, die die Daten nachnutzen, darüber informiert sind.³⁰

30 Vgl. Hannah Schwarz, „Data consistency,“ in *Data Processing and Documentation: Generating High Quality Research Data in Quantitative Social Science Research* (GESIS Papers), hrsg. von Sebastian Netscher und Christina Eder (Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, in Vorbereitung).

4.2. Disziplinspezifische strukturierte Metadaten für Auffindbarkeit und Zitierbarkeit

Wie oben geschildert, wird den Forschungsdaten mit der sogenannten „Dokumentation“ relevante Kontextinformation in unstrukturierter Form zur Seite gestellt. Diese gewährleistet die Interpretierbarkeit der Daten, ist aber aufgrund der mangelnden Maschinenlesbarkeit nur bedingt geeignet, die Auffindbarkeit der Daten einerseits und die schnelle Beurteilung der Relevanz der Daten für die eigene Forschung andererseits zu ermöglichen. Diese Funktion erfüllen strukturierte Metadaten, die unter Verwendung disziplinspezifischer Standards erstellt werden – von den datengebenden Wissenschaftlerinnen und Wissenschaftlern selbst bei Einreichung und/oder von den Kuratorinnen und Kuratoren, die die Daten dann ins Archiv übernehmen. Das GESIS Datenarchiv für Sozialwissenschaften verwendet dafür ein eigenes Metadatenschema.³¹ Es bestehen Mappings zu disziplinspezifischen Standards, insbesondere dem *dalra* Metadatenschema für die Registrierung von DOIs in den Sozial- und Wirtschaftswissenschaften sowie den XML-basierten Schemata DDI-Codebook und DDI-Lifecycle der Data Documentation Initiative (<http://www.ddialliance.org>).³² Diese ermöglichen eine differenzierte Beschreibung der Forschungsdaten, einschließlich der Erhebung und Aufbereitung, und machen es damit einfacher, Forschungsdaten für die Nutzung in der eigenen Forschung zu finden und deren Relevanz zu beurteilen. Die vom GESIS Datenarchiv für Sozialwissenschaften vorgenommene Registrierung der Forschungsdaten bei *dalra*, der DOI-Registrierungsagentur für Sozial- und Wirtschaftsdaten, sichert darüber hinaus auch die eindeutige Identifizierbarkeit und Zitierbarkeit der Daten.

4.3. Ausgewählte Unterstützungsangebote

GESIS unterstützt Forschende, die mit sozialwissenschaftlichen Methoden arbeiten, beim Management und der langfristigen Sicherung und Bereitstellung ihrer Forschungsdaten. Dies geschieht über ein Schulungs- und Beratungsangebot sowie über die Bereitstellung einer Infrastruktur für die Veröffentlichung und Archivierung von Daten. So führt GESIS einführende Workshops in Themen des Forschungsdatenmanagements für Sozialwissenschaftlerinnen und wissenschaftler durch und unterstützt bei Fragen zur informierten Einwilligung oder der Pseudonymisierung von quantitativen Forschungsdaten.³³

Im Bereich der unterstützenden Infrastruktur bietet das GESIS Datenarchiv für Sozialwissenschaften mit dem Repository datorium eine Lösung an, die es Forschenden ermöglicht, ihre Forschungsdaten und zugehörige Dokumentation eigenständig mit Metadaten zu beschreiben und anderen unter standardisierten Lizenzbedingungen zur Nutzung zur Verfügung zu stellen. Zur Unterstützung von Institutionen, die selbst kein eigenes (disziplinspezifisches) Repository aufsetzen wollen, bietet das GESIS Datenarchiv für Sozialwissenschaften *SowiDataNet* als kostenpflichtige Business-to-business-Lösung

31 Wolfgang Zenk-Möltgen und Norma Habel, *Der GESIS Datenbestandskatalog und sein Metadatenschema. Version 1.8*, GESIS Technical Reports, Nr. 1 (2012), <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-292372>.

32 Ute Koch et al., *dalra Metadata Schema Documentation for the Publication and Citation of Social and Economic Data Version 4.0*, GESIS Papers, Nr. 25 (2017), <https://doi.org/10.4232/10.mdsdoc.4.0>; Esra Akdeniz und Wolfgang Zenk-Möltgen, *DDI-Lifecycle im Datenarchiv: das Metadatenschema für die Dokumentation in verschiedenen Softwaresystemen*, GESIS Papers, Nr. 2 (2017), <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-50354-9>.

33 <https://www.gesis.org/angebot/>.

an. Bei beiden Lösungen steht das GESIS Datenarchiv den Forschenden und kuratierenden Institutsmitgliedern als Ansprechpartner bei inhaltlichen und technischen Fragen zur Verfügung (siehe Tabelle 4). Weitere Angebote des GESIS Datenarchivs für Sozialwissenschaften umfassen z.B. die Langzeitarchivierung (u.a. mit Migration der Daten in andere Dateiformate) und die added-value Aufbereitung von Forschungsdaten (z.B. Erschließung auf Variableneben).

Tabelle 4: Überblick datorium und SowiDataNet

	datorium	SowiDataNet
Zielgruppe	Kleine Projekte, Einzelforscher/innen	Forschungseinrichtungen ohne eigenes Repository
Workflow	<ul style="list-style-type: none"> • Beschreibung mit Metadaten und Upload der Dateien durch die Forschenden; • Überprüfung von Daten und Dokumentation durch GESIS Datenkurator/in; • DOI-Registrierung und Veröffentlichung. 	<ul style="list-style-type: none"> • Beschreibung mit Metadaten und Upload der Dateien durch Institutsmitglieder (Forschende oder Kurator/Kuratorin); • Zweistufige Überprüfung der Daten durch Institution und GESIS Datenkurator/in; • DOI-Registrierung und Veröffentlichung.
Archivierung	Bitstream Preservation, d.h. Dateien werden im ursprünglichen Format langfristig erhalten	
URL	https://datorium.gesis.org/xmlui/	http://www.sowidatanet.de/xmlui

5. Sicherung und Bereitstellung qualitativer Forschungsdaten

Das GESIS Datenarchiv für Sozialwissenschaften bietet einen umfassenden Service für die Archivierung und Bereitstellung quantitativer Forschungsdaten. Für die weniger standardisierten qualitativen Forschungsdaten existiert bisher ein solch umfassendes Angebot noch nicht, obwohl auch hier die Nachnutzung der Daten eine wichtige Rolle spielt.³⁴ Daher wird in jüngster Zeit der Ruf nach einer Forschungsinfrastruktur auch für qualitative Daten laut.³⁵ Zwei themenspezifische Angebote existie-

34 Vgl. Denis Huschka und Claudia Oellers, „Einführung: Warum qualitative Daten und ihre Sekundäranalyse wichtig sind,“ in *Forschungsinfrastrukturen für die qualitative Sozialforschung*, hrsg. Denis Huschka et al. (Berlin: Scivero, 2013), 9-18, zuletzt geprüft am 14.02.2018, https://www.germandataforum.org/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf.

35 Vgl. Hubert Knoblauch, „Einigen Anforderungen an forschungsinfrastrukturen aus der Sicht der qualitativen Forschung,“ in *Forschungsinfrastrukturen für die qualitative Sozialforschung*, hrsg. Denis Huschka et al. (Berlin: Scivero, 2013), 27-34, zuletzt geprüft am 14.02.2018, https://www.germandataforum.org/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf.

ren bereits, die hier erwähnt werden sollen. Zum einen unterstützt der Qualiservice der Universität Bremen,³⁶ ausgehend vom Archiv für Lebenslaufforschung, die Archivierung von qualitativen Interviewstudien. Im Bereich der Bildungsforschung archiviert das FDZ Bildung am Deutschen Institut für Internationale Pädagogische Forschung³⁷ qualitative Beobachtungs- und Interviewdaten und stellt diese in anonymisierter Form für die wissenschaftliche Nachnutzung zur Verfügung.

6. Schlussfolgerungen

Zur Sicherung der guten wissenschaftlichen Praxis sollen Wissenschaftlerinnen und Wissenschaftler den Forschungsprozess möglichst transparent gestalten und die erhobenen Daten und erzielten Ergebnisse reproduzierbar und nachvollziehbar machen. Im Sinne von „Open Science“ sollen sie zudem eine möglichst freie und breite Nachnutzung ihrer Daten und Ergebnisse ermöglichen. Dies stellt Forschende aber auch vor große Hürden, da erhobene Daten möglichst für eine Nachnutzung dokumentiert und bereitgestellt werden sollen und gleichzeitig die Persönlichkeitsrechte der Teilnehmenden geschützt werden müssen.

Um Forschende hierbei zu unterstützen, bietet das GESIS-Datenarchiv für Sozialwissenschaften umfassende Informations- und Serviceangebote rund um die Archivierung quantitativer Forschungsdaten. Im Bereich der qualitativen Sozialforschung wurden mit dem Angebot von Qualiservice und dem FDZ Bildung bereits erste Grundsteine für eine unterstützende Infrastruktur gelegt. Jedoch muss sichergestellt werden, dass diese Angebote die Forschenden auch erreichen. Vor diesem Hintergrund erscheint es sinnvoll, die bislang noch nicht umfassend ausgeprägte Verzahnung generischer, insbesondere hochschulischer Angebote mit disziplinspezifischen Diensten zu stärken. Im großen Maßstab spielen diese Überlegungen nicht zuletzt in den aktuellen Planungen zur Nationalen Forschungsdateninfrastruktur (NFDI) eine Rolle, die sich wesentlich dadurch auszeichnen soll, dass generische und disziplinspezifische Angebote sich ergänzen, um unnötige Redundanzen zu vermeiden.³⁸ Im „Kleinen“ kann dies bedeuten, die Kooperation von Hochschulen und disziplinspezifischen Angeboten beim Management, der Bereitstellung und der langfristigen Sicherung von Forschungsdaten zu intensivieren – etwa durch Vereinbarungen zur Veröffentlichung von Daten in bestimmten fachlichen Repositorien, die Einbindung von externer Expertise in Lehrveranstaltungen (z.B. im Rahmen von Gastvorträgen) oder auch bei der Beratung von Forschenden zu Spezialthemen wie dem Datenschutz.

36 <http://www.qualiservice.org/>.

37 Vgl. Doris Bambey und Marc Rittberger, „Das Forschungsdatenzentrum (FDZ) Bildung des DIPF: Qualitative Daten der empirischen Bildungsforschung im Kontext.“ in *Forschungsinfrastrukturen für die qualitative Sozialforschung*, hrsg. von Denis Huschka et al., 2013, 63-72, zuletzt geprüft am 14.02.2018, https://www.germandataforum.org/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf.

38 <https://www.fdz-bildung.de>. Vgl. Rat für Informationsinfrastrukturen, *Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*, 2016, zuletzt geprüft am 14.02.2018, <http://www.rfii.de/?p=1998>.

Literaturverzeichnis

- Akdeniz, Esra und Wolfgang Zenk-Möltgen. *DDI-Lifecycle im Datenarchiv: das Metadatenchema für die Dokumentation in verschiedenen Softwaresystemen*. GESIS Papers, Nr. 2 (2017). <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-50354-9>.
- Arbeitsgruppe Datenschutz und qualitative Sozialforschung Working. *Datenschutzrechtliche Anforderungen bei der Generierung und Archivierung qualitativer Interviewdaten*. Working Paper Series des Rates für Sozial- und Wirtschaftsdaten, Nr. 238 (2014). <http://hdl.handle.net/10419/97181>.
- Article 29 Data Protection Working Party. „Opinion 05/2014 on Anonymisation Techniques.“ 0829/14/EN WP216 (2014). Zuletzt geprüft am 16.02.2018. http://collections.internetmemory.org/haeu/20140424090503/http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.
- Bambey, Doris und Marc Rittberger. „Das Forschungsdatenzentrum (FDZ) Bildung des DIPF: Qualitative Daten der empirischen Bildungsforschung im Kontext.“ In *Forschungsinfrastrukturen für die qualitative Sozialforschung*, herausgegeben von Denis Huschka, Hubert Knoblauch, Claudia Oellers und Heike Solga, 63-72, 2013. Zuletzt geprüft am 14.02.2018. https://www.germandataforum.org/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf.
- Bundesministerium für Bildung und Forschung (BMBF). „Bekanntmachung Richtlinie zur Förderung von Forschung zu ‚Digitalisierung im Bildungsbereich – Grundsatzfragen und Gelingensbedingungen‘.“ Bundesanzeiger AT 26.09.2017 B2 (2017). Zuletzt geprüft am 14.02.2018. <https://www.bmbf.de/foerderungen/bekanntmachung-1420.html>.
- „bwFDM - Koordiniertes Forschungsdatenmanagement in Baden-Württemberg.“ Zuletzt geprüft am 14.02.2018. <https://bwfdm.scc.kit.edu/>.
- Cabrera, Noemi Betancort und Elmar Haake. *Das Qualiservice Metadatenchema. Version 1.0*. Qualiservice Technical Reports, Nr. 13 (2013). <http://nbn-resolving.de/urn:nbn:de:gbv:46-00103310-19>.
- Curdt, Constanze, Volker Hess, Ania Lopez, Benedikt Magrean, Dominik Rudolph und Johanna Vompras. „Herausforderung Forschungsdatenmanagement – Unterstützung der Hochschulen durch eine einrichtungübergreifende Kooperation in NRW.“ In *E-Science-Tage 2017: Forschungsdaten managen*, herausgegeben von Jonas Kratzke und Vincent Heuveline. Heidelberg: heiBOOKS, 2017. <https://doi.org/10.11588/heibooks.285.377>.

- Deutsche Forschungsgemeinschaft. *Sicherung guter wissenschaftlicher Praxis / Safeguarding Good Scientific Practice. Denkschrift / Memorandum*, 2013. Zuletzt geprüft am 10.03.2018. http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf.
- Deutsche Forschungsgemeinschaft. *Umgang mit Forschungsdaten: DFG-Leitlinien zum Umgang mit Forschungsdaten*, 2015. Zuletzt geprüft am 27.06.2018. http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/.
- Deutsche Gesellschaft für Soziologie und Berufsverband Deutscher Soziologinnen und Soziologen. „Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS)“, 2017. Zuletzt geprüft am 10.03.2018. <http://www.soziologie.de/de/die-dgs/ethik/ethik-kodex.html>.
- Eisentraut, Marcus. „Data anonymization.“ In *Data Processing and Documentation: Generating High Quality Research Data in Quantitative Social Science Research* (GESIS Papers), herausgegeben von Sebastian Netscher und Christina Eder. Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, in Vorbereitung.
- Europäische Menschenrechtskonvention (Konvention zum Schutze der Menschenrechte und Grundfreiheiten) vom 04.11.1950. Zuletzt geprüft am 14.02.2018. <https://dejure.org/gesetze/MRK>.
- European Commission. *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*. 2016. Zuletzt geprüft am 14.02.2018. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- Harzenetter, Karoline. „Metadata documentation at the variable level.“ In *Data Processing and Documentation: Generating High Quality Research Data in Quantitative Social Science Research* (GESIS Papers), herausgegeben von Sebastian Netscher und Christina Eder. Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, in Vorbereitung.
- Huschka, Denis, und Claudia Oellers. „Einführung: Warum qualitative Daten und ihre Sekundäranalyse wichtig sind.“ In *Forschungsinfrastrukturen für die qualitative Sozialforschung*, herausgegeben von Denis Huschka, Hubert Knoblauch, Claudia Oellers und Heike Solga, 9-18. Berlin: Scivero, 2013. Zuletzt geprüft am 14.02.2018. https://www.germandataforum.org/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf.
- Jakobsson, Ulf, Ricarda Braukmann und Malin Lundgren. „Expert tour guide on Data Management. Chapter 1: Plan.“ 2017. Zuletzt geprüft am 14.02.2018. <http://www.cessda.eu/DMGuide>.

- Jensen, Uwe. *Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten*. GESIS Technical Reports, Nr. 7 (2012). <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-320650>.
- Knoblauch, Hubert. „Einigen Anforderungen an Forschungsinfrastrukturen aus der Sicht der qualitativen Forschung“. In *Forschungsinfrastrukturen für die qualitative Sozialforschung*, herausgegeben von Denis Huschka, Hubert Knoblauch, Claudia Oellers und Heike Solga, 27-34. Berlin: Scivero, 2013. Zuletzt geprüft am 14.02.2018. https://www.germandataforum.org/dl/downloads/forschungsinfrastrukturen_qualitative_sozialforschung.pdf.
- Koch, Ute, Esra Akdeniz, Jana Meichsner, Brigitte Hausstein und Karoline Harzenetter. *dalra Metadata Schema Documentation for the Publication and Citation of Social and Economic Data Version 4.0*. GESIS Papers, Nr. 25 (2017). <https://doi.org/10.4232/10.mdsdoc.4.0>.
- Meyermann, Alexia und Maike Porzelt. *Hinweise zur Anonymisierung von qualitativen Daten*. *forschungsdaten bildung informiert*, Nr. 1 (2014). Zuletzt geprüft am 14.02.2018. https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert-nr-1.pdf.
- Rat für Informationsinfrastrukturen. *Leistung aus Vielfalt: Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*, 2016. Zuletzt geprüft am 14.02.2018. <http://www.rfii.de/?p=1998>.
- Rat für Sozial- und Wirtschaftsdaten. *Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften. Orientierungshilfen für die Beantragung und Begutachtung datengenerierender und datennutzender Forschungsprojekte*. RatSWD Output Series, Nr. 3 (2016). <https://doi.org/10.17620/02671.7>.
- Schaar, Katrin. *Was hat die Wissenschaft beim Datenschutz künftig zu beachten? Allgemeine und spezifische Änderungen beim Datenschutz im Wissenschaftsbereich durch die neue Europäische Datenschutzgrundverordnung*. Working Paper Series des Rates für Sozial- und Wirtschaftsdaten, Nr. 257 (2016). <https://doi.org/10.17620/02671.19>.
- Schwarz, Hannah. „Data consistency.“ In *Data Processing and Documentation: Generating High Quality Research Data in Quantitative Social Science Research* (GESIS Papers), herausgegeben von Sebastian Netscher und Christina Eder. Köln: GESIS - Leibniz-Institut für Sozialwissenschaften, in Vorbereitung.
- Verbund Forschungsdaten Bildung. *Hinweise zur Erstellung eines Methoden- / Feldberichts*. *fdbinfo*, Nr. 3 (2015). Zuletzt geprüft am 14.02.2018. https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdbinfo_3.pdf.

- 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 (DSG) (2016). Zuletzt geprüft am 14.02.2018. <http://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679&from=DE>.
- Watteler, Oliver. *Erstellung von Methodenberichten für die Archivierung von Forschungsdaten*, 2010. Zuletzt geprüft am 14.02.2018. https://www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Aufbau_Methodenbericht_v1_2010-07.pdf.
- Watteler, Oliver und Katharina E. Kinder-Kurlanda. „Anonymisierung und sicherer Umgang mit Forschungsdaten in der empirischen Sozialforschung.“ *Datenschutz und Datensicherheit: DuD; Recht und Sicherheit in Informationsverarbeitung und Kommunikation* 39, Nr. 8 (2015): 515-19. <https://doi.org/10.1007/s11623-015-0462-0>.
- Wirth, Heike. *Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific-Use-File*, 2006. Zuletzt geprüft am 14.02.2018. <https://www.destatis.de/DE/Methoden/Methodenpapiere/Mikrozensus/Arbeitspapiere/Arbeitspapier11.html>.
- Wirth, Heike. „Die faktische Anonymität von Mikrodaten: Ergebnisse und Konsequenzen eines Forschungsprojektes.“ *ZUMA Nachrichten* 16, Nr. 30 (1992): 7-65. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-209679>.
- Wissenschaftsrat. *Bericht zur wissenschaftsgeleiteten Bewertung umfangreicher Forschungsinfrastrukturvorhaben für die Nationale Roadmap*, 2017. Zuletzt geprüft am 14.02.2018. <https://www.wissenschaftsrat.de/download/archiv/6410-17.pdf>.
- Zenk-Möltgen, Wolfgang und Norma Habel. *Der GESIS Datenbestandskatalog und sein Metadatenchema. Version 1.8*. GESIS Technical Reports, Nr. 1 (2012). Zuletzt geprüft am 14.02.2018. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-292372>.