



Reihe: FAIR-Prinzipien und Datenschutz.

Wie kann FAIRes Datenmanagement mit der Datenschutzgrundverordnung vereinbart werden?

Factsheet 3

Data Linkage: FAIR und datenschutzkonform

Autor:innen:

Antonia Schier¹, Oliver Vettermann², Oliver Watteler³, Ulrich Sax⁴

Affiliations:

- 1) KonsortSWD, Leibniz-Institut für Bildungsverläufe (LIbBi), Bamberg, DE [bis Sommer 2024]
- 2) NFDI4Culture, FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, Karlsruhe, DE
- 3) KonsortSWD, GESIS - Leibniz-Institut für Sozialwissenschaften, Köln, DE
- 4) NFDI4Health, Universitätsmedizin Göttingen, Institut für Medizinische Informatik, Göttingen, DE

Stand: 01.07.2025

Dieses Factsheet ist Teil einer Reihe zum Themenbereich "FAIR-Prinzipien und Datenschutz". Diese Factsheet-Reihe ist in einer Arbeitsgruppe zum Thema Datenschutz im Rahmen der [Nationalen Forschungsdateninfrastruktur](#) entstanden.

Die Factsheets verfolgen ausschließlich informative Zwecke und ersetzen keine individuelle Rechtsberatung.

Lizenz:



CC BY 4.0 – <https://creativecommons.org/licenses/by/4.0/>

INHALT

Data Linkage – FAIR und datenschutzkonform	2
Data Linkage – Einführung und Herausforderungen	2
Forschungsdaten von Personen und Data Linkage	3
Fallbeispiele ‚Data Linkage‘	4
Survey Daten + Regionaldaten	4
Survey Daten + qualitative Befragung	5
Fallbeispiele ‚Record Linkage‘	5
Verlinkung von Survey und Social Media Daten	5
Verlinkung Survey Daten und administrative Daten	6
Verlinkung von klinischen Daten und genetischen Daten	6
Weiterführende Literatur	7
Querverweise	9
Quellen	9

Data Linkage – FAIR und datenschutzkonform

Das vorliegende Fact Sheet führt kurz in den Themenkomplex Data Linkage und die Herausforderungen ein, bietet Definitionen zentraler Konzepte und stellt anhand von Fallbeispielen kurz mögliche Lösungen verschiedener Formen von Data Linkage dar. Ein abschließender Abschnitt verweist auf Literatur, in denen der praktische Umgang mit Data Linkage dargestellt wird und die als weiterführende Wegweiser für die eigene Arbeit dienen können.

Data Linkage – Einführung und Herausforderungen

Durch die Anwendung der FAIR-Prinzipien (für das Originalkonzept siehe Wilkinson et al., 2016; zur Umsetzung siehe das Webangebot der [GoFAIR-Initiative](#)) sollen Forschung und die Nutzung von Forschungsdaten transparenter und effizienter werden, da z.B. doppelte Datenerhebungen vermieden werden. Aus verbesserten und leichteren Zugängen zu Forschungsdaten ergeben sich neue Forschungsfragen. Herausforderungen entstehen in diesem Zusammenhang durch einen zunehmenden Bedarf an der Verknüpfung verschiedener Datenbestände in der wissenschaftlichen Gemeinschaft, dem ‚Data Linkage‘ oder ‚Record Linkage‘. Verknüpft werden etwa einzelne Datensätze aus verschiedenen Forschungsprojekten oder Forschungsdaten mit administrativen, Gesundheits- oder Regionaldaten.

Solche Verlinkungen bergen Risiken bezüglich des Datenschutzes, da z.B. Daten von Individuen aus verschiedenen Quellen miteinander kombiniert werden und man durch die kombinierten Datenpunkte genauere Angaben zu den Individuen erhalten kann. Dies kann etwa bei Verlinkungen von Umfragedaten

Definition: ‚Data Linkage‘

‚Data Linkage‘ steht generell für die Verknüpfung von Daten aus verschiedenen Quellen mit dem Ziel der gemeinsamen Analyse (Beuthner et al., 2021).

Definition: ‚Record Linkage‘

Ein spezieller Fall des ‚Data Linkage‘ ist das ‚Record Linkage‘, welches sich auf die Verknüpfung von Daten derselben Personen aus unterschiedlichen Datenbasen oder -beständen bezieht (Schnell, 2016).

mit Sozialversicherungsdaten der Fall sein, oder bei einer ähnlichen Verlinkung von Umfragen mit Social-Media-Daten. Die kombinierten Daten müssen dabei noch keinen direkten Personenbezug erkennen lassen, aber durch die Kombination nähern sie sich den Individuen quasi an. Die durch die Kombination möglicherweise entstehenden Risiken müssen mit den Anliegen von FAIR nach verbessertem Datenzugang in Beziehung gesetzt werden und datenschutzkonform ablaufen.

Um Verlinkungen datenschutzkonform durchführen zu können, müssen Rechtsgrundlagen für die Datenverarbeitung existieren, die Zusammenführungen verschiedener Bestände ermöglichen. Ferner müssen technische und organisatorische Maßnahmen – kurz TOMS – getroffen werden, um Risiken für betroffene Personen zu minimieren. Diese Bedingungen können einem möglichst offenen Zugang zu solchen kombinierten Forschungsdaten im Sinne der FAIR-Kriterien entgegenstehen, da man sie z.B. nicht

einfach mit einer offenen Lizenz zum Download bereitstellen kann. So müssen die verlinkten Daten entweder in ihrem Informationsgehalt reduziert werden, nur nach Abschluss weiterer Datennutzungsvereinbarungen zugänglich gemacht werden, oder der Zugriff muss bspw. auf Analysen in sicheren Arbeitsumgebungen (Secure- oder Safe-Rooms) beschränkt werden (Watteler 2022, 241-243).

Forschungsdaten von Personen und Data Linkage

Forschungsdaten sind allgemein „Daten, die im Zuge wissenschaftlicher Vorhaben entstehen, z. B. durch Beobachtungen, Experimente, Simulationsrechnungen, Erhebungen, Befragungen, Quellenforschungen, Aufzeichnungen, Digitalisierung, Auswertungen“ (Rat für Informationsinfrastrukturen, 2016). Viele wissenschaftliche Fachbereiche wie die Sozial-, Medizin- oder Verhaltensforschung befassen sich mit Daten von und über Menschen (sog. Human Subject Research). Diese Daten werden in vielfältiger Weise u.a. in Forschungsprojekten erhoben und im Anschluss häufig in pseudonymisierter oder anonymisierter Weise über Repositorien, Forschungsdatenzentren oder Datenarchive für weitere Vorhaben angeboten.

Datenschutzrechtlich haben wir es in oben genannten Projekten regelmäßig mit persönlichen Daten zu tun, solange z.B. Kontaktinformationen wie Adressen noch Teil des Datenbestandes sind. Der Umgang mit solchen Daten bedarf der besonderen Sorgfalt und über Jahrzehnte wurden Arbeitsabläufe etabliert, die im Einklang mit Datenschutzrecht und Forschungsethik eine langfristige Sicherung dieser Daten und eine Publikation zur Nachnutzung ermöglichen. Beispielhaft sind hier die Sozialwissenschaften (Rat Für Sozial- und Wirtschaftsdaten (RatSWD), 2020) und die medizinische Forschung (Pommerening et al., 2014).

Von Personen erhobene Forschungsdaten können eventuell mit anderen Daten derselben Personen („Record Linkage“) oder mit Kontextdaten verknüpft werden („Data Linkage“). Hierzu müssen keine direkten Identifikatoren wie Namen und Adressen mehr in den Daten enthalten sein. Solche Verknüpfungen bergen Risiken bezüglich des Datenschutzes, da die einzelnen Datensätze vielleicht anonymisiert sind, aber in Kombination doch Rückschlüsse auf die Personen zulassen. Derartige Risiken besprechen wir in den Fallbeispielen.

Um Daten aus verschiedenen Quellen miteinander verlinken zu können, bedarf es einer eigenen Rechtsgrundlage. Eine Möglichkeit ist das Einholen informierter Einwilligungen siehe auch [FACTSHEET Einwilligung](#)), welches frühzeitig im Projektverlauf eingeplant werden sollte. Die Rechtsgrundlage wird durch technische Verfahren, wie die Verknüpfung der Daten über Pseudonyme und entsprechende ID-Prozeduren ergänzt. Oft erfordert das Verlinken auch eine datenschutzrechtliche Abstimmung und vertragliche Vereinbarungen zwischen den verantwortlichen Institutionen, deren Daten genutzt werden sollen.

Das Zusammenführen von ursprünglich personenbezogenen Daten, die also nun anonymisiert oder pseudonymisiert vorliegen, birgt dabei immer das potentielle Risiko einer Re-Identifizierung. Pseudonymisierte Daten unterliegen dem Datenschutz, weil die Daten als Pseudonym wie ein Zeiger für

Definition: Gemeinsame Verantwortlichkeit

Mehrere datenverarbeitende Stellen sind nach Art. 26 DSGVO gemeinsam verantwortlich, wenn sie sich über die beabsichtigten Zwecke und die bei der Verarbeitung verwendeten Mittel geeinigt haben. Dabei kann jede Stelle individuelle Zwecke und Ziele bei der Verarbeitung verfolgen; auch eine Hierarchie der Beteiligten ist möglich. Wichtig ist die gemeinsame und damit abhängige Entscheidungsfindung.

zumindest eine beteiligte Person/Institution wieder aufdeckbar sind. Anonymisierte Daten verfügen über ein geringes Risiko, weshalb sie nicht mehr dem Datenschutzrecht unterliegen. Die zunehmende Bearbeitung von Datensätzen in und mit Machine-Learning-Algorithmen zur Mustererkennung erhöht die Risiken allerdings, da durch nach einer Anonymisierung/Pseudonymisierung verbleibende Fragmente in den zusammengeführten Daten auch eine

Identifizierung möglicher wird. Die verschiedenen kombinierten Risiken einzelner Datensätze sind deshalb dynamisch zu begreifen. Um Menschen vor Risiken zu schützen, sind Datenverarbeitung mit ausreichenden Schutzmaßnahmen zu begrenzen. Auf einem Spektrum von Maßnahmen und Risiken reichen die Formen von anonymen Daten bis zu personenbezogenen Daten und besonderen personenbezogenen Daten mit zwingenden notwendigen Schutzmaßnahmen.

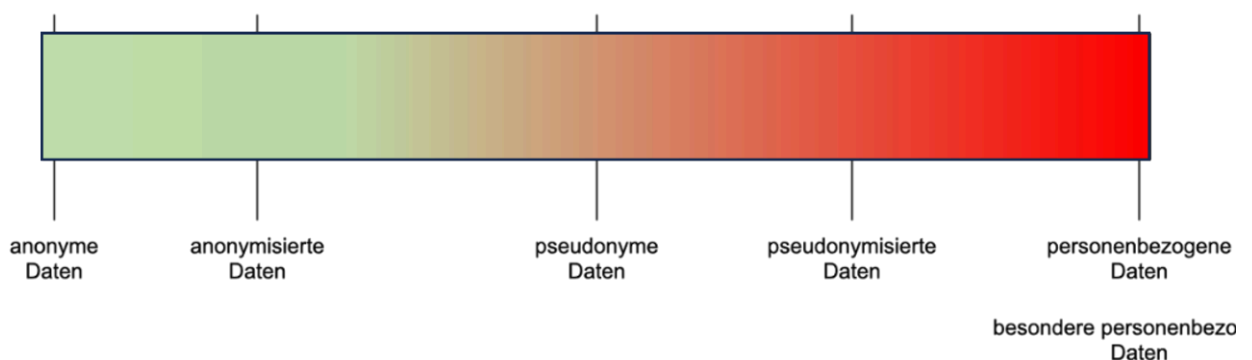


Abbildung: Verschiedene Grade des Personenbezugs (eigene Darstellung)

Fallbeispiele ‚Data Linkage‘

Survey Daten + Regionaldaten

Ein Bereich des Data Linking ist das Kontextualisieren von Survey-Daten durch Daten der Regionalstatistik. Dieser Bereich der Statistik ist ein „Instrument, das dazu dient, raumrelevante Informationen in allen sozialen und ökonomischen Bereichen zu gewinnen und regionale Strukturzusammenhänge, Verflechtungsbeziehungen und Entwicklungsvorgänge zu analysieren (Boustedt, 1975)“. Raumrelevante Informationen können auf verschiedenen politisch-administrativen Ebenen wie Bundesländern, Regierungsbezirken und Kreisen vorliegen.

Für die Verknüpfung von regional-statistischen Daten mit Surveys bedarf es sog. Georeferenzen. Als georeferenzierte Survey-Daten werden oft solche Daten bezeichnet, die standardisierte Raumbezüge wie

Namen und Bezeichner (ID) für Kreise, Gemeinden oder Postleitzahlengebiete umfassen. Durch kleinräumige Verortung von Befragten sollen die Kontexte sozialen Handelns besser erfasst und verstanden werden. Beispiele für solches Handeln sind politisches Verhalten und Einstellungen, soziale Bedingungen von Gesundheit, oder sozial-räumliche Einflüsse auf Bildungsübergänge (Müller, 2019).

Das Risiko bei dieser Art des Linkings besteht darin, dass der räumliche Kontext wie etwa ein administrativer Kreis eindeutig über statistische Angaben identifizierbar ist. Enthält nun der Survey Angaben zu Personen wie Geschlecht, genaue Berufsangaben und Alter, kann es sein, dass diese Angaben auf nationaler Ebene unproblematisch sind, aber auf regionaler Ebene schnell einem kleinen Personenkreis zugeordnet werden können. Hier kann z.B. mit der Kategorisierung von regionalen Angaben gearbeitet werden, die keine Rückschlüsse z.B. auf die genannten Kreise mehr zulassen.

Survey Daten + qualitative Befragung

In der sozialwissenschaftlichen Forschung kommt vermehrt der Bedarf auf, bestehende quantitative Daten groß angelegter Panel-Surveys gezielt mit qualitativen Daten kleinerer Studien und Forschungsprojekte zu verknüpfen (Mixed-Methods-Studien, (Creswell & Plano Clark, 2011). Hierdurch wird das Forschungspotential der Daten erweitert (Legewie & Tucci, 2016).

Die Verknüpfung erfordert eine informierte Einwilligung der Befragten, die im Voraus bei der Studienplanung mitgedacht, eingeholt und dokumentiert (Zentrum Für Studienmanagement – Arbeitsbereich Datenschutz in Erhebungen, Leibniz-Institut für Bildungsverläufe (LifBi), 2023)) werden muss. Häufig führen mehrere Forschungsinstitute diese unterschiedlichen Studien durch und arbeiten ggf. zusammen. Je nachdem wie viele Institute dabei personenbezogene Daten in welcher Art und Weise verarbeiteten, erfordert dies ggf. einen datenschutzrechtlichen Vertrag zur Auftragsverarbeitung oder gemeinsamen Verantwortung, bei denen u.a. die technischen und organisatorischen Maßnahmen festgelegt und die Wahrung der Informationspflichten den Teilnehmenden gegenüber sowie deren Betroffenenrechte geregelt werden.

Durch die unterschiedliche und sich ergänzende Informationsdichte der qualitativen mit den quantitativen Daten erhöht sich gleichzeitig das Re-Identifizierungspotential der befragten Personen. Somit erhöhen sich auch die Anforderungen an die Anonymisierung der Forschungsdaten für die Verarbeitung und die Bereitstellung für die wissenschaftliche Gemeinschaft.

Fallbeispiele ‚Record Linkage‘

Verlinkung von Survey und Social Media Daten

Social Media Daten können als Daten definiert werden, die durch Internet-basierte Applikationen generiert werden und oft sog. User generated content enthalten. Dieser „Content“ meint, durch die

Plattformnutzenden erzeugte Inhalte wie Texte, Bilder, Videos oder Tonaufnahmen (Obar & Wildman, 2015). Solche Daten haben den Vorteil, dass sie die Untersuchung digitalen menschlichen Verhaltens auf eine Art und Weise ermöglichen, die den betroffenen Personen nicht unmittelbar bewusst ist ('non intrusive'). Gleichzeitig bilden die Daten das genannte Verhalten teilweise mit hoher Präzision und Granularität ab (sog. „Digital Trace“ oder “Digital Behavioral Data”).

Viele Studien, die sich ausschließlich auf derartige Daten stützen, können den Nachteil haben, dass relevante Informationen über andere Aktivitäten (z.B. Verhalten in politischen Wahlen) sowie Charakteristika der Untersuchungspersonen (z.B. Bildungsabschlüsse oder Einkommen) fehlen. Diese Daten können durch Surveys gewonnen und mit den digitalen Verhaltensdaten verknüpft werden (Stier et al., 2020).

Das Risiko bei dieser Art des Linkings besteht darin, dass „user generated content“ wieder den ursprünglichen Nutzerkonten und damit weiteren Inhalten und ggf. persönlichen Daten zugeordnet werden kann. Damit ergeben sich ähnliche Szenarien wie im Falle der Kontextdaten. Vermeintlich anonyme Survey-Daten könnten nach einer Verlinkung mit Social Media Daten wieder den ursprünglichen Befragten zugeordnet werden. Eine Re-Identifizierung würde zumindest erleichtert. Es bedarf also auch bei dieser Art von Verlinkung einer Rechtsgrundlage wie einer informierten Einwilligung und die kombinierten Survey und Social Media Daten können nur unter geschützten Bedingungen zur Analyse angeboten werden.

Verlinkung Survey Daten und administrative Daten

Ein weiterer Bereich, in dem Survey-Daten als Teil verknüpfter Datenbestände genutzt werden, ist deren Verbindung mit administrativen Daten. In öffentlichen Verwaltungen, wie z.B. bei der Sozialversicherung, fallen laufend sog. prozess-produzierte Daten an, die auch als Registerdaten bezeichnet werden. Die Vorteile dieser Daten im Vergleich zu Daten aus Befragungen liegen in ihrer Vollständigkeit und Genauigkeit. Außerdem entlasten sie die untersuchten Personen und ihre Nachnutzung senkt die Kosten für eine eigene Datenbeschaffung.

Bei der Verknüpfung von prozess-produzierten Daten mit beispielsweise Survey-Daten können nach Schnell (2016) neben dem 'Record Linkage' die Erzeugung von synthetischen Dateien aus verschiedenen Datenquellen („Data fusion“), sowie die Verknüpfung von Individualdaten mit Aggregatdaten („Enhancing data“) unterschieden werden. Personenbezogene Daten, die zu administrativen Zwecken erhoben werden, haben in der Regel eine eigene Rechtsgrundlage (z.B. die Datenerfassungs- und -Übermittlungsverordnung – DEÜV – für Angaben zur Sozialversicherung auf der Basis des Sozialgesetzbuches). Sollen diese Daten mit Survey-Daten verknüpft werden, bedarf es einer neuen Rechtsgrundlage, die dies erlaubt.

Die Risiken für diese Fallgruppe werden anhand eines Beispiels erläutert. Die Befragungsdaten des Nationalen Bildungspanels (NEPS) wird mit administrativen Daten des Instituts für Berufsforschung (IAB), die auf den Sozialversicherungsdaten der Bundesagentur für Arbeit basieren, verknüpft

(„NEPS-ADIAB“-Datensatz). Zentral für die Verknüpfung der beiden Datensätze sind die Adress- und Kontaktdaten der NEPS-Teilnehmenden sowie deren Geburtsdatum und Geschlecht. Die Routinen basieren auf etablierten Prozeduren, die durch das „German Record Linkage Center“ für den deutschen Kontext angepasst wurden (Antoni & Schnell, 2019). Der NEPS-ADIAB-Datensatz wird regelmäßig aktualisiert, erweitert und ist über sichere Vor-Ort- und Remote-Zugänge in den zuständigen Forschungsdatenzentren für die wissenschaftliche Gemeinschaft zugänglich.

Verlinkung von klinischen Daten und genetischen Daten

In vielen klinischen Studien werden ergänzend zur Erhebung der klinischen Daten zu verschiedenen Zeitpunkten auch Biomarker untersucht. In der Regel wird dazu Biomaterial entnommen, das dann sequenziert wird. Die sequenzierten Daten müssen im Sinne eines genetischen Reports ausgewertet und interpretiert werden.

Eine Verknüpfung wäre aus zwei Blickwinkeln interessant: (1) Ergänzung einer klinischen Datenerhebung mit Biomarkern bzw. genetischer Varianten und Mutationen (aggregierte Daten); (2) Annotation einer Sammlung von Biomaterial oder Sequenzierungsrohdaten mit klinischen Daten.

Aus Sicht der Risiken einer Re-Identifizierung ist der Fall 1 nicht signifikant komplexer zu sehen als die vorhergehenden Fälle mit der Einschränkung, dass das Berichten von seltenen Mutationen oder einer Kombination vieler Varianten per se ein eindeutiges Muster pro untersuchtem Material darstellen können. Für den Fall, dass eine solche Kombination von Varianten und Mutationen an einer Stelle mit einer Identität versehen sein sollte, könnten weitere Vorkommen dieser personenbezogenen Daten ermittelt werden. Im Fall 2 sind bei der Nutzung der Sequenzierungsrohdaten nicht nur die Rechte der Personen betroffen, von denen das Material gewonnen wurde, sondern auch die aller Verwandten nullten (Zwilling), ersten (Eltern bzw. Kinder) und zweiten Grades (Geschwister). Daher ist ergänzend zur DSGVO ([Art. 9 Abs. 1 lit. g](#) bzw. [Abs. 4 DSGVO](#), [Erwägungsgrund 34 DSGVO](#)) das [Gesetz über genetische Untersuchungen bei Menschen \(Gendiagnostikgesetz – GenDG\)](#) anzuwenden. D.h. vor allem, dass Ergebnisse genetischer Untersuchungen und Analysen nach einer Schutzfrist zu löschen sind, soweit keine schutzwürdigen Interessen der betroffenen Person beeinträchtigt werden ([§ 12 GenDG](#)).

Weiterführende Literatur

Das Themenfeld Data Linkage ist in den vergangenen Jahren enorm angewachsen. Neue Gesetzgebung und technische Möglichkeiten haben auch Chancen durch neue Verknüpfungsmöglichkeiten in verschiedenen Disziplinen eröffnet. Wie die Beispiele zeigen, lässt sich diese Verarbeitung auch mit den FAIR-Prinzipien und rechtlichen Vorgaben vereinbaren. In der Regel werden bestehende Risiken minimiert (siehe Beispiel zu Regionaldaten) oder das Realisieren des Risikos durch technische und organisatorische Maßnahmen aktiv verhindert (siehe Beispiel zu administrativen Daten). Die folgenden Arbeiten zielen darauf ab,

Verlinkungen verschiedener Art für die Forschungspraxis handhabbar zu machen und leisten auf diesem Wege Beiträge, um verknüpfte Daten FAIR anzubieten.

Zentral sind die Arbeiten von Reiner Schnell zum Bereich Data Linkage, die im Text mehrfach zitiert werden. Einen zusammenfassenden Überblick bietet Schnell (2020). Mit einem Schwerpunkt auf Data Linkage in der medizinischen Forschung haben March et al., (2019) einen Überblick erstellt, der auch eine Checkliste zur Bewertung geplanter Forschungsvorhaben bietet. Das NFDI-Konsortium NFDI4Health hat ein umfassendes White Paper zur Verbesserung der Record Linkage für die Gesundheitsforschung in Deutschland erstellt, welches zehn Use Cases umfasst und auch auf genetische Daten eingeht (Intemann et al., 2023). Schließlich geben Beuthner et al. (2021) weiterführende Hinweise zur Verlinkung von sozialwissenschaftlichen Umfragen mit räumlichen, Social-Media- sowie Sensordaten. Mögliche Maßnahmen zur Risikominimierung enthält das Standard-Datenschutzmodell des ULD Schleswig-Holstein (Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder, 2024).

Quellen

- Antoni, M., & Schnell, R. (2019). The Past, Present and Future of the German Record Linkage Center (GRLC). *Jahrbücher Für Nationalökonomie Und Statistik*, 239(2), 319–331. <https://doi.org/10.1515/jbnst-2017-1004>
- Beuthner, C., Breuer, J., & Jünger, S. (2021). *Data Linking—Linking survey data with geospatial, social media, and sensor data (GESIS Survey Guidelines)* (Version 1.0). GESIS - Leibniz Institute for the Social Sciences. https://doi.org/10.15465/GESIS-SG_EN_039
- Boustedt, O. (1975). Grundriß der empirischen Regionalforschung. Schroedel.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed). SAGE Publications.
- Intemann, T., Kaulke, K., Kipker, D.-K., Lettieri, V., Stallmann, C., Schmidt, C. O., Geidel, L., Bialke, M., Hampf, C., Stahl, D., Lablans, M., Franke, M., Kraywinkel, K. P., Kieschke, J., Bartholomäus, S., Näher, A.-F., Temper, G., Lambarki, M., March, S., ... Semler, S. C. (2023). *White Paper—Verbesserung des Record Linkage für die Gesundheitsforschung in Deutschland* (S. 167). <https://repository.publisso.de/resource/frl:6461898>
- Legewie, N., & Tucci, I. (2016). *Panel-basierte Mixed- Methods-Studien. Design, Feldzugang, Potentiale und Herausforderungen am Beispiel der Studie „Das Erwachsenwerden der Nachkommen von GastarbeiterInnen in Deutschland“*. SOEPpapers 872. <https://hal.science/hal-01394134>
- March, S., Andrich, S., Drepper, J., Horenkamp-Sonntag, D., Icks, A., Ihle, P., Kieschke, J., Kollhorst, B., Maier, B., Meyer, I., Müller, G., Ohlmeier, C., Peschke, D., Richter, A., Rosenbusch, M.-L., Scholten, N., Schulz, M., Stallmann, C., Swart, E., ... Hoffmann, F. (2019). Gute Praxis Datenlinkage (GPD). *Das Gesundheitswesen*, 81(08/09), 636–650. <https://doi.org/10.1055/a-0962-9933>
- Müller, S. (2019). Räumliche Verknüpfung georeferenzierter Umfragedaten mit Geodaten: Chancen, Herausforderungen und praktische Empfehlungen. In U. Jensen, S. Netscher, & K. Weller (Eds.), *Forschungsdatenmanagement sozialwissenschaftlicher Umfragedaten* (pp. 211–229). Verlag Barbara Budrich. <https://doi.org/10.3224/84742233.13>
- Obar, J. A., & Wildman, S. (2015). Social media definition and the governance challenge: An introduction to the special issue. *Telecommunications Policy*, 39(9), 745–750. <https://doi.org/10.1016/j.telpol.2015.07.014>
- Pommerening, K., Drepper, J., Helbing, K., & Ganslandt, T. (2014). *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten: Generische Lösungen der TMF 2.0*. Medizinisch Wissenschaftliche Verlagsgesellschaft. <https://doi.org/10.32745/9783954662951>
- Rat für Informationsinfrastrukturen. (2016). Bericht des Redaktionsausschusses Begriffe an den RfII (Nr.1; RfII Berichte). <https://rfii.de/?p=2039>
- Rat für Sozial- und Wirtschaftsdaten (RatSWD). (2020). Handreichung Datenschutz, 2. Überarbeitete Auflage. *RatSWD Output Paper Series*. <https://doi.org/10.17620/02671.50>
- Schnell, R. (2016). Record Linkage. In C. Wolf, D. Joye, T. W. Smith, & Y. Fu (Hrsg.), *The SAGE Handbook of Survey Methodology*.
- Schnell, R. (2020). Record Linkage als zentraler Baustein der Forschung mit Registern und Big Data-Nutzungen. In B. Klumpe, J. Schröder, & M. Zwick (Eds.), *Qualität bei zusammengeführten Daten: Befragungsdaten, administrative Daten, neue digitale Daten: Miteinander besser?* (pp. 147–162). Springer Fachmedien. https://doi.org/10.1007/978-3-658-31009-7_11

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, 38(5), 503–516.

<https://doi.org/10.1177/0894439319843669>

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>

Zentrum für Studienmanagement – Arbeitsbereich Datenschutz in Erhebungen, Leibniz-Institut für Bildungsverläufe (LifBi). (2023). Wie holt man eine gültige informierte Einwilligung ein und wie wird diese dokumentiert? <https://doi.org/10.5157/LIFBI:DS:FS01:1.0>