

Gute Praxis Datenlinkage

Stefanie March¹, Silke Andrich^{2,3}, Johannes Drepper⁴, Dirk Horenkamp-Sonntag⁵,
Andrea Icks^{2,3}, Peter Ihle⁶, Joachim Kieschke⁷, Bianca Kollhorst⁸, Birga Maier⁹, Ingo
Meyer⁶, Gabriele Müller¹⁰, Christoph Ohlmeier¹¹, Dirk Peschke^{12,13}, Adrian Richter¹⁴,
Marie-Luise Rosenbusch¹⁵, Nadine Scholten¹⁶, Mandy Schulz¹⁵, Christoph Stallmann¹,
Enno Swart¹, Stefanie Wobbe-Ribinski¹⁷, Antke Wolter¹⁷, Jan Zeidler¹⁸, Falk
Hoffmann¹⁹

¹ Institut für Sozialmedizin und Gesundheitssystemforschung (ISMG), Medizinische Fakultät, O.-v-G.-Universität
Magdeburg, Magdeburg

²⁻¹⁹ siehe Abstract



MEDIZINISCHE FAKULTÄT
UNIVERSITÄTSKLINIKUM MAGDEBURG A.Ö.R.

Gliederung

1. Hintergrund und Begriffsbestimmung
2. Ausgangssituation in Deutschland
3. Quo Vadis Datenlinkage in Deutschland
4. Gute Praxis Datenlinkage
5. Diskussion



Hintergrund und Begriffsbestimmung

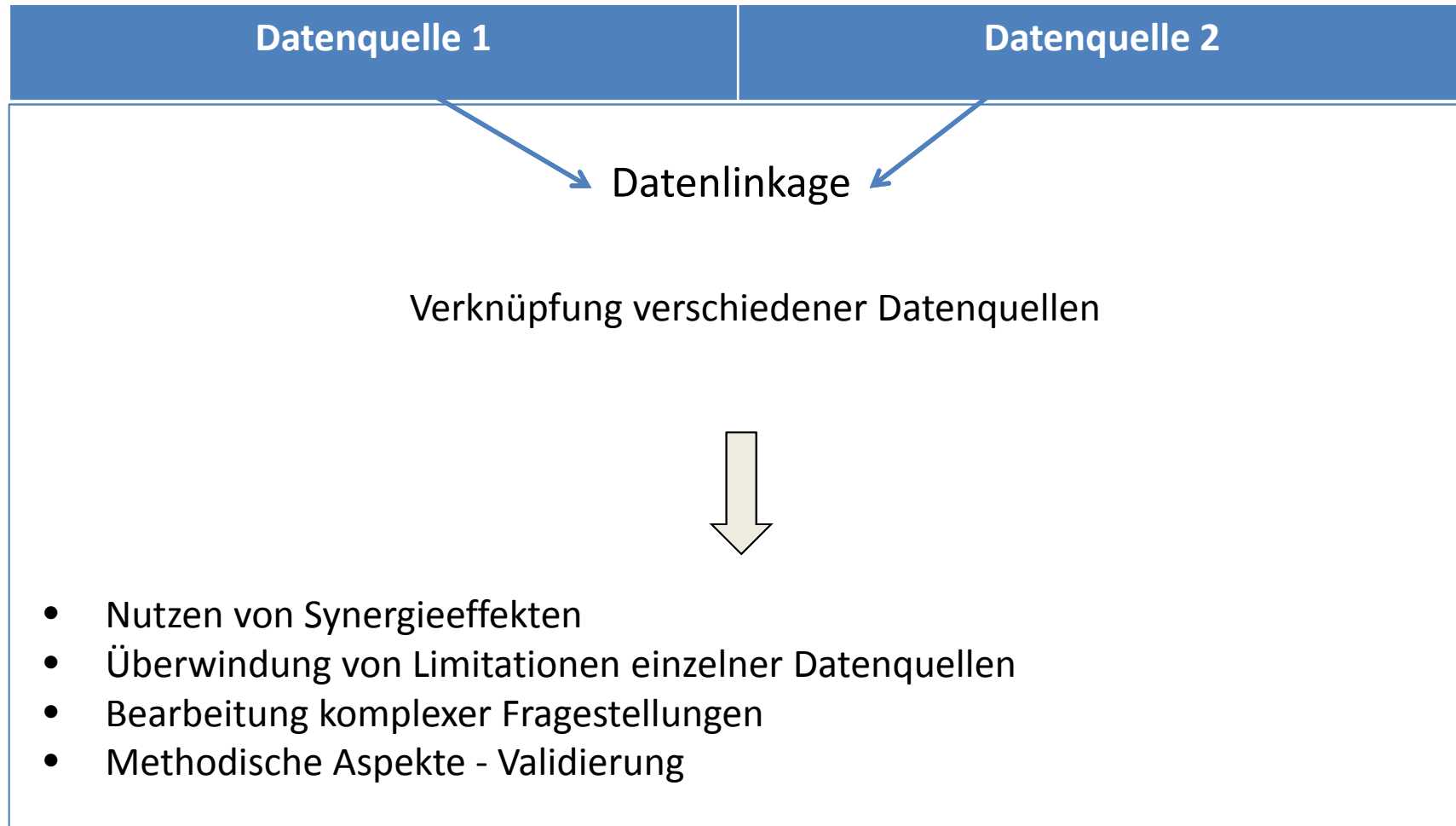
Gesundheits- und Sozialforschung: Verwendung einer Vielzahl an verschiedenen Datenquellen

⇒ Unterscheidung in Primär- und Sekundärdaten

- **Primärdaten** (z. B. Geyer et al. 2003, Hoffmann et al. 2008b, Ohlmeier et al. 2014, Swart et al. 2015)
 - Erhebung und Analyse zu bestimmten wissenschaftlichen Zwecken mit konkreter Fragestellung
 - regionale oder bundesweite Erhebungen
 - Beispiele: Gesundheitsstudien des Robert Koch-Instituts (RKI)
- **Sekundärdaten** (z. B. Swart und Ihle 2005, Swart et al. 2014, Swart et al. 2015)
 - Verwendung über den ursprünglichen Erhebungszweck hinaus
 - Beispiel: Daten der Gesetzlichen Krankenkassen (GKV) = administrative Daten der jeweiligen gesetzlichen Krankenkasse (auch „Routinedaten“)



Kombination verschiedener Datenquellen



Ausgangssituation in Deutschland

- Im internationalen Vergleich eine recht junge Methodik
 - Bisher sehr spezifische Erfahrungen zum Datenlinkage, z. B. Krebsregistrierung (z. B. Kajüter et al. 2014)
 - Forderung nach deutschen Datenlinkage-Studien im Rahmen der Gesundheits- und Versorgungsforschung (z. B. Hoffmann et al. 2008, Icks et al. 2010, Bitzer 2015, Horenkamp-Sonntag et al. 2016)
 - Aktuell: Ausschreibung Innovationsausschuss des Gemeinsamen Bundesausschusses (G-BA) vom 19.10.2018 ⇒ Förderbekanntmachung Versorgungsforschung zum themenoffenen Bereich:
 - (iii) Aufbereitung und Verknüpfung von Gesundheitsdaten aus verschiedenen Quellen zur Verbesserung der Patientenversorgung
- ⇒ Es fehlt derzeit an Handreichungen für Forschende, die ein Datenlinkage planen.
- ⇒ Autarke Forschungsgruppen mit Neuerarbeitung in aufwändigem Prozess; bislang wenig Unterstützung (z.B. German Record Linkage Center)



Projektgruppe „Gute Praxis Datenlinkage“ (Konstituierung Februar 2016)

Erstellung Manuskript in 2016/2017

Das Gesundheitswesen

DOI: 10.1055/s-0043-125070

eFirst <https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0043-125070>, Zugriff am 19.02.2019

eFirst Ausgabe Meistgelesen

Publikationsdatum: 20 Februar 2018

Übersichtsarbeit

March, Stefanie; Antoni, Manfred; Kieschke, Joachim; Kollhorst, Bianca; Maier, Birga; Müller, Gabriele; Sariyar, Murat; Schulz, Mandy; Enno, Swart; Zeidler, Jan; Hoffmann, Falk:
Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme
Quo Vadis Data Linkage in Germany? An Initial Inventory

Gliederung Quo Vadis Datenlinkage (March et al. 2018)

1. Verfahren und Arten des Datenlinkage

- Exaktes und fehlertolerantes Linkage
- Regelbasiertes, distanzbasiertes und probabilistisches Linkage
- Linkage mit direkten und indirekten Identifikatoren
- Linkage mit Klartextangaben und mit verschlüsselten Identifikatoren
- Linkage mit und ohne Einwilligungserklärung
- Linkage mit unterschiedlichen Formen des Blockings

2. Rechtlicher Rahmen und Datenschutz

- Welche Datenquellen sollen verknüpft werden?
- Was ist bei personenidentifizierenden Merkmalen zu beachten?
- Wann brauche ich eine Einwilligungserklärung für ein Datenlinkage?

3. Softwaretools

4. Qualitätskontrollen

- Überprüfung auf mögliche Fehler
- Plausibilitätskontrollen
- Qualität der Identifikatoren
- Datenlinkage im zeitlichen Verlauf



Quo Vadis Datenlinkage - Beispiele für verschiedene Arten des Datenlinkage

(Auszug)

Studientitel	Welche Daten wurden verknüpft?	Welche Arten des Datenlinkage wurden verwendet?
lidA - leben in der Arbeit - eine Kohortenstudie zu Gesundheit und Älterwerden in der Arbeit (March et al. 2012, 2015; www.lidA-studie.de)	Befragungsdaten mit Daten der Gesetzlichen Krankenversicherung (GKV) und der Bundesagentur für Arbeit (BA)	deterministisches Linkage, Linkage mit direkten Identifikatoren, Linkage mit Einwilligungserklärung
	Befragungsdaten mit aggregierten Daten der GKV	Linkage mit indirekten Identifikatoren, Linkage ohne Einwilligungserklärung (Antrag nach § 75 SGB X)
Vergleich unterschiedlicher Linkageverfahren sowie Vollständigkeit klinischer Angaben (Ohlmeier et al. 2015)	Daten der GKV mit Daten eines Krankenhausinformationssystems	Linkage mit direkten und indirekten Identifikatoren
QS-AMI Studie (Maier et al. 2015)	Daten der GKV mit Daten eines klinischen Registers	deterministisches Linkage, Linkage mit indirekten Identifikatoren

(March et al. 2018, e25)



Quo Vadis Datenlinkage

- Erste Bestandsaufnahme im Kontext der deutschen Rahmenbedingungen
- Praxisnahe Überarbeitung und in strukturierter Form durch die Projektgruppe
- Kollegialer Austausch

Gute Praxis Datenlinkage

- Mai 2018: Kick-off-Treffen
- Erweiterung Projektgruppe auf 23 Mitglieder
- Februar 2019: Manuskript zur Kommentierung in den Arbeitsgruppen
 - Erhebung und Nutzung von Sekundärdaten (AGENS) der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Epidemiologie (DGEpi)
 - Validierung und Linkage von Sekundärdaten des Deutschen Netzwerks Versorgungsforschung (DNVF)
 - Datenschutz sowie IT-Infrastruktur und Qualitätsmanagement der TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V.





Jahr

2019

Ausgabe

07: 525-574 | e110-e140

06: 457-518

05: 377-452 | e109

04: 293-369 | e93-e108

03: 161-224 | e34-e92 | De

02: 80-156

01: 5-71 | e1-e33

- › Inhaltsverzeichnis
- › aktuelle Ausgabe
- › Probeausgabe (01/2019)



CC BY-NC-ND 4.0 · Gesundheitswesen
DOI: 10.1055/a-0962-9933



Originalarbeit

Eigentümer und Copyright ©Georg Thieme Verlag KG 2018

Gute Praxis Datenlinkage (GPD)

Good Practice Data Linkage

Stefanie March, Silke Andrich, Johannes Drepper, Dirk Horenkamp-Sonntag, Andrea Icks, Peter Ihle, Joachim Kieschke, Bianca Kollhorst, Birga Maier, Ingo Meyer, Gabriele Müller, Christoph Ohlmeier, Dirk Peschke, Adrian Richter, Marie-Luise Rosenbusch, Nadine Scholten, Mandy Schulz, Christoph Stallmann, Enno Swart, Stefanie Wobbe-Ribinski, Antke Wolter, Jan Zeidler, Falk Hoffmann

› [Institutsangaben](#)

› [Weitere Informationen](#)

› Auch verfügbar auf **eRef**

Abstract Volltext Referenzen Abbildungen

› [als PDF herunterladen](#)

- › Zusammenfassung
- › Abstract
- › Ziele und Zielgruppe der GPD
- › Anwendungsfelder
- › Methodik
- › Leitlinie 1: Forschungsziele, Fragestellungen, Datenquellen und Ressourcen
 - › Empfehlung 1.1: Forschungsziele, Fragestellungen und ggf. Hypothesen müssen so präzise wie möglich formuliert werden, um darauf aufbauend ein detailliertes Anforderungsprofil für die zu nutzenden Daten zu erstellen
 - › Empfehlung 1.2: Es muss frühzeitig geprüft werden, ob die zur Verfolgung der Forschungsziele und Beantwortung der Forschungsfragen notwendigen Informationen verfügbar sind. Darüber hinaus müssen das Forschungsdesign, relevante Beobachtungszeiträume und die zu untersuchende Population spezifiziert werden
 - › Empfehlung 1.3: Unter dem Aspekt Datenherkunft (Data provenance) müssen die Dateneigner Entwicklungen der Datenquellen offenlegen und kommunizieren. Insbesondere sind historische/systematische Veränderungen an den Datenquellen, wie die Einführung neuer Kodierungssysteme oder von Schlüsselvariablen, zu kommunizieren

Ziel der Projektgruppe (March et al. 2019)

„Mit der Guten Praxis Datenlinkage (GPD) wird ein **Standard für die Durchführung von Forschungsvorhaben in der Gesundheits- und Sozialforschung** formuliert, die ein Datenlinkage **personenbezogener Daten** nach wissenschaftlichen Grundsätzen anstreben.“ (ebd., S. 637)

Zielsetzungen

1. Handlungsanleitung im Sinne einer Leitlinie für Wissenschaftlerinnen und Wissenschaftler, die ein Forschungsvorhaben mit Datenlinkage planen und durchführen
2. Checkliste Bewertung geplanter Vorhaben und Publikationen über solche



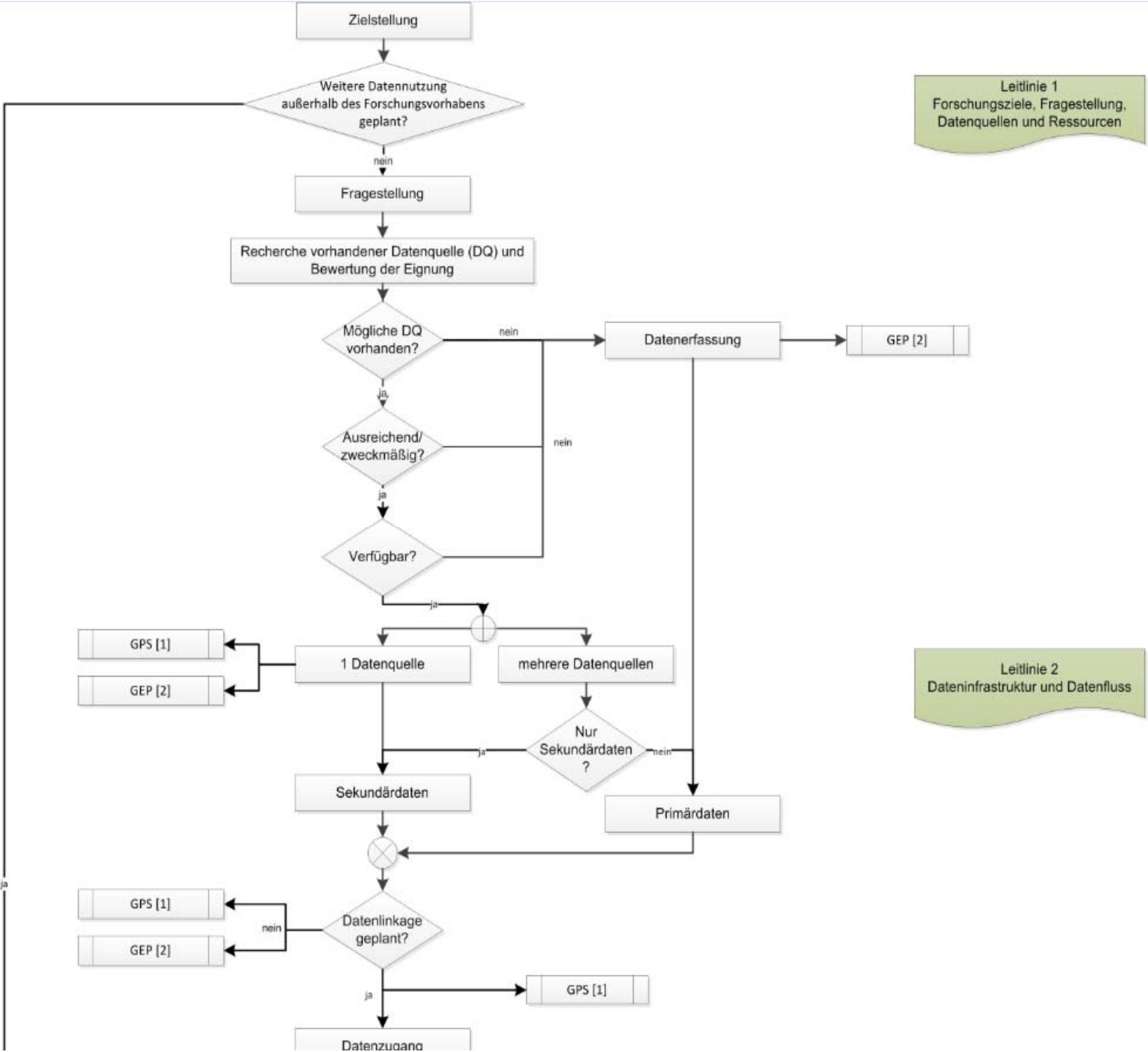
Leitlinien der GPD (March et al. 2019)

1. Forschungsziele, Fragestellung, Datenquellen und Ressourcen
2. Dateninfrastruktur und Datenfluss
3. Datenschutz
4. Ethik
5. Schlüsselvariablen und Linkageverfahren
6. Datenprüfung/ Qualitätssicherung
7. Langfristige Datennutzung für noch festzulegende Fragestellungen

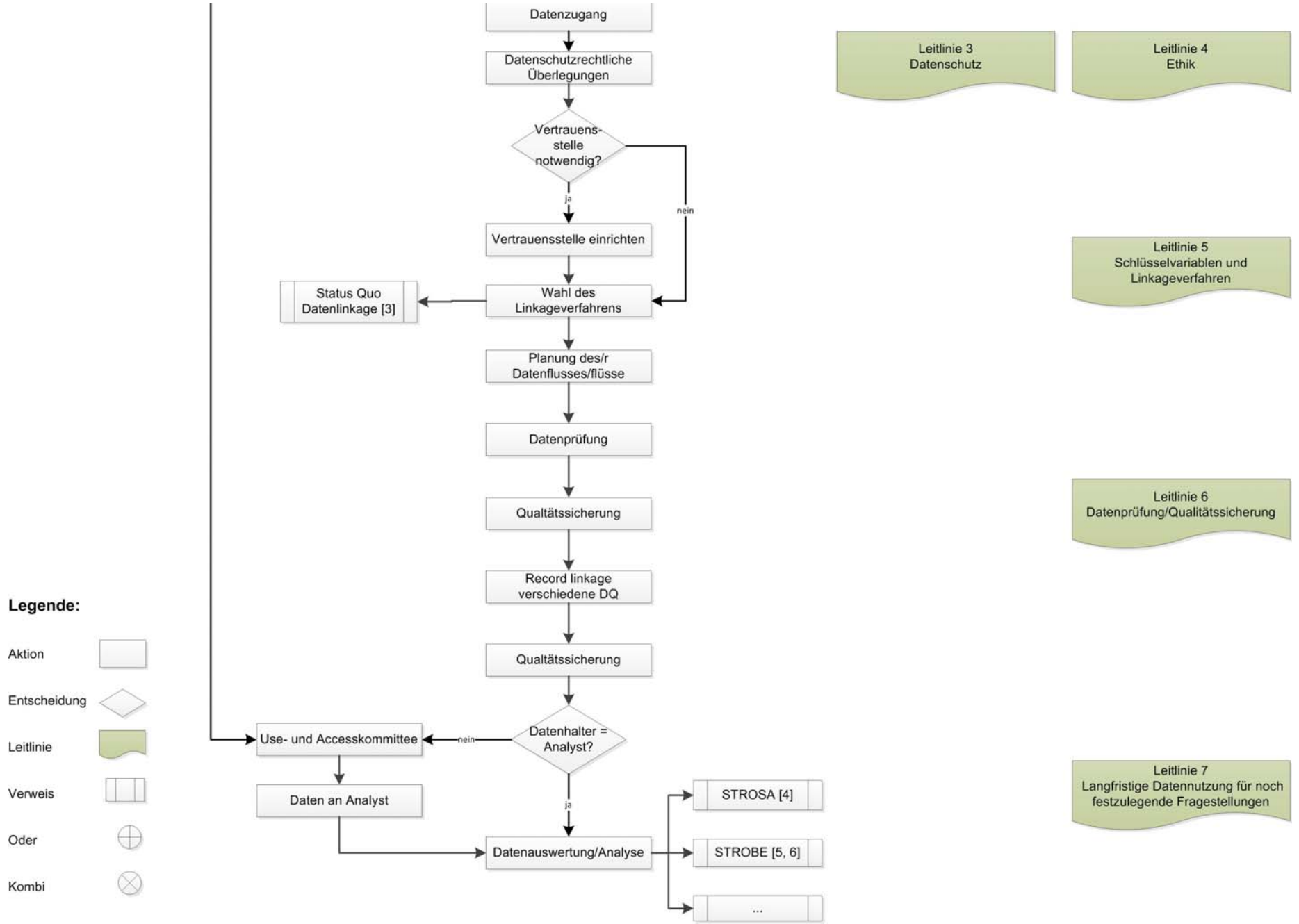


je Leitlinie verschiedene Empfehlungen

Strukturiertes, prozessorientiertes Inhaltsverzeichnis der GPD (I) (modifiziert nach March et al. 2019, S. 641)



Strukturiertes, prozessorientiertes Inhaltsverzeichnis der GPD (II) (modifiziert nach March et al. 2019, S. 641)



Beispiel Leitlinie 2 - Dateninfrastruktur und Datenfluss (March et al. 2019)

Empfehlungen:

- 2.1 Der Datenfluss und die Verantwortlichkeiten müssen eindeutig festgelegt werden.
- 2.2 Die allgemeinen technischen und organisatorischen Anforderungen an die Datenübermittlung (siehe Leitlinie 6.1 der GPS) müssen beachtet und Besonderheiten bei Vorhaben mit Datenlinkage berücksichtigt werden.
- 2.3 Je nach Linkageverfahren muss eine geeignete Software eingesetzt werden.
- 2.4 Für die Löschung von Daten und das Widerspruchsmanagement muss ein geeigneter Prozess definiert werden.



Beispiel Leitlinie 3 – Datenschutz (March et al. 2019)

Empfehlungen:

- 3.1 Datenschutzrechtliche Vorgaben müssen bereits in der Planung und bis zum Abschluss des Vorhabens berücksichtigt und eine Deanonymisierung/ Reidentifizierung einzelner Personen durch das Linkage verhindert werden.
- 3.2 Es muss geprüft werden, ob eine Einwilligungserklärung notwendig ist.
- 3.3 Ein Datenschutzkonzept muss erarbeitet werden.
- 3.4 Wird ein Linkage erst im Nachhinein in einem Forschungsvorhaben geplant, müssen die sich ggf. ergebenden datenschutzrechtlichen Vorgaben sorgfältig geprüft werden.



Abschließende Diskussionsfragen

- Gibt es Anmerkungen zum vorgestellten Stand?



Kontakt:

: stefanie.march@med.ovgu.de

: +49 391 6724323

Literatur auf Anfrage

Vielen Dank für Ihre Aufmerksamkeit!



Stefanie March¹, Silke Andrich^{2,3}, Johannes Drepper⁴, Dirk Horenkamp-Sonntag⁵, Andrea Icks^{2,3}, Peter Ihle⁶, Joachim Kieschke⁷, Bianca Kollhorst⁸, Birga Maier⁹, Ingo Meyer⁶, Gabriele Müller¹⁰, Christoph Ohlmeier¹¹, Dirk Peschke^{12,13}, Adrian Richter¹⁴, Marie-Luise Rosenbusch¹⁵, Nadine Scholten¹⁶, Mandy Schulz¹⁵, Christoph Stallmann¹, Enno Swart¹, Stefanie Wobbe-Ribinski¹⁷, Antke Wolter¹⁷, Jan Zeidler¹⁸, Falk Hoffmann¹⁹

¹ Institut für Sozialmedizin und Gesundheitssystemforschung (ISMG), Medizinische Fakultät, Otto-von-Guericke-Universität Magdeburg, Magdeburg

² Institut für Versorgungsforschung und Gesundheitsökonomie, Centre for Health and Society, Medizinische Fakultät, Heinrich-Heine-Universität Düsseldorf, Düsseldorf

³ Institut für Versorgungsforschung und Gesundheitsökonomie, Deutsches Diabetes-Zentrum (DDZ), Leibniz-Zentrum für Diabetes-Forschung an der Heinrich-Heine-Universität Düsseldorf, Düsseldorf

⁴ TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V., Berlin

⁵ Techniker Krankenkasse, Hamburg

⁶ PMV forschungsgruppe, Köln

⁷ Epidemiologisches Krebsregister Niedersachsen, Registerstelle, Oldenburg

⁸ Abteilung Biometrie und EDV, Leibniz-Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen

⁹ Berlin-Brandenburger Herzinfarktregister e. V., Berlin

¹⁰ Zentrum für Evidenzbasierte Gesundheitsversorgung (ZEGV), Universitätsklinikum und Medizinische Fakultät Carl Gustav Carus, TU Dresden, Dresden

¹¹ IGES Institut GmbH, Berlin

¹² Institut für Public Health und Pflegeforschung (IPP), Universität Bremen, Bremen

¹³ Department für Angewandte Gesundheitswissenschaften, Studienbereich Physiotherapie, Hochschule für Gesundheit Bochum, Bochum

¹⁴ Institut für Community Medicine, Universitätsmedizin Greifswald, Greifswald

¹⁵ Zentralinstitut für die kassenärztliche Versorgung in Deutschland (Zi), Fachbereich Versorgungsforschung, Systemanalyse und Data Science, Berlin

¹⁶ Institut für Medizinsoziologie, Versorgungsforschung und Rehabilitationswissenschaft (IMVR), Humanwissenschaftliche Fakultät und Medizinische Fakultät, Universität zu Köln, Köln

¹⁷ DAK Gesundheit, Vorstandsreferat Versorgungsforschung, Hamburg

¹⁸ Center for Health Economics Research Hannover (CHERH), Leibniz Universität Hannover, Hannover

¹⁹ Department für Versorgungsforschung, Fakultät für Medizin und Gesundheitswissenschaft, Carl von Ossietzky Universität Oldenburg, Oldenburg

