

Heisig, Jan Paul; Li, Jianghong; Allmendinger, Jutta

**Book Part — Published Version**

## Gesundheitsdaten als öffentliches Gut

**Provided in Cooperation with:**

WZB Berlin Social Science Center

*Suggested Citation:* Heisig, Jan Paul; Li, Jianghong; Allmendinger, Jutta (2021) : Gesundheitsdaten als öffentliches Gut, In: van der Eijk, Philip Ganten, Detlev Marek, Roman (Ed.): Was ist Gesundheit? Interdisziplinäre Perspektiven aus Medizin, Geschichte und Kultur, ISBN 978-3-110-71333-6, De Gruyter, Warsaw, Berlin, pp. 363-375, <https://doi.org/10.1515/9783110713336-025>

This Version is available at:

<https://hdl.handle.net/10419/247425>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

# Jan Paul Heisig, Jianghong Li und Jutta Allmendinger

## Gesundheitsdaten als öffentliches Gut

**Abstract:** *Health Data as a Public Good.* Routine health data, which are collected by health insurers and other agencies in the health care system, offer enormous potential for health monitoring and research. Germany has been slow to make such data available for socially beneficial purposes, partly due to concerns about privacy and data protection. Against this background, we discuss some of the most important potential uses of routine health data and call for a broader societal debate about the benefits, risks, and appropriate regulation of routine health data usage. We then review the Western Australian Data Linkage System as an example of a data infrastructure that is characterized by high levels of stakeholder and patient involvement and a sophisticated method of privacy protection. While Germany does not need to copy this approach, we hope that the experiences of Western Australia and other countries will stimulate and inform the overdue debate about a modern, responsible, and sustainable approach to socially beneficial health data usage in Germany.

Sollte der Titel dieser Ausführungen ein Fragezeichen tragen? Diese Frage wurde immer wieder an uns herangetragen. Unsere deutliche Antwort: Nein. Gesundheitsdaten müssen zumindest teilweise als ein öffentliches Gut verstanden und behandelt werden. Ein öffentliches Gut sind Gesundheitsdaten, da ihre sachgerechte und angemessene Verwendung zu Forschungs-, Beobachtungs- und Berichtszwecken erheblichen gesellschaftlichen Nutzen mit sich bringen kann und deshalb grundsätzlich möglich sein sollte.

Mit „öffentlich“ ist dabei aber keinesfalls „frei und für alle zugänglich“ gemeint. Ganz im Gegenteil: Bei Gesundheitsdaten handelt es sich um höchst sensible Informationen, deren Nutzung besonders strengen Regeln unterworfen werden sollte.


Wenn wir in diesem Beitrag von „Gesundheitsdaten“ sprechen, verweisen wir vor allem auf Daten, die im Zuge der alltäglichen Routinen im Gesundheitssystem „quasi nebenbei“ anfallen. Wichtige Beispiele sind die Diagnose-, Behandlungs- und Abrechnungsdaten, die bei niedergelassenen Ärztinnen und Ärzten und in


---

**Jan Paul Heisig**, Wissenschaftszentrum Berlin für Sozialforschung; Freie Universität Berlin

**Jianghong Li**, Wissenschaftszentrum Berlin für Sozialforschung

**Jutta Allmendinger**, Wissenschaftszentrum Berlin für Sozialforschung; Humboldt-Universität zu Berlin

 OpenAccess. © 2021 Jan Paul Heisig, Jianghong Li und Jutta Allmendinger, publiziert von De Gruyter.

 Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 4.0 International Lizenz.

<https://doi.org/10.1515/9783110713336-025>

Krankenhäusern entstehen. Solche Daten sind also prozessproduziert und werden auch als „Routinedaten“ bezeichnet. Primäre Zwecke der Datenerhebung und Dokumentation sind die optimale Behandlung der Patient\*innen und die Erfassung und Abrechnung von Gesundheitsdienstleistungen. Die Generierung von Daten für die Forschung und Gesundheitsberichterstattung ist bei ihrer Entstehung allenfalls ein untergeordnetes Ziel.

Aufgrund der fortschreitenden Digitalisierung von Arbeits- und Verwaltungsprozessen liegen solche Routinedaten zunehmend in digitaler Form vor, was ganz neue Chancen und Risiken eröffnet: Chancen, weil verschiedene Datenquellen ohne großen Aufwand verknüpft werden können, neue und vor wenigen Jahrzehnten noch undenkbare Formen der Datennutzung möglich werden und sich der Kreis der Datennutzer\*innen fast ohne zusätzliche Kosten erweitern lässt. Die Kehrseite dieser Chancen liegt in Risiken, die sich aus dem Zusammenfügen von Datenquellen und dem daraus entstehenden „Buch des Lebens“ ergeben, welches höchst sensible Daten von vor der Geburt bis hin zum Tod enthält (Dunn 1946; Holman et al. 2008). Wird dieses Buch unzureichend abgesichert, entstehen verheerende Formen des Datenmissbrauchs oder -diebstahls.

Dieses Nebeneinander von Chancen und Risiken bedeutet, dass allen Entscheidungen bezüglich der Nutzung von Gesundheitsdaten ein Abwägungsprozess zugrunde liegen muss, der die individuellen und kollektiven Vorteile den individuellen und kollektiven Risiken einer Nutzung gegenüberstellt. Diese Abwägung darf nicht hinter verschlossenen Türen stattfinden. Eine breite gesamtgesellschaftliche Diskussion, klare und transparente Verfahrensregeln, eine offene Kommunikation von Entscheidungen sowie angemessene Widerspruchs- und Beschwerdemöglichkeiten sind für die gesellschaftliche Akzeptanz einer gemeinwohlorientierten Nutzung von Gesundheitsdaten unerlässlich.

Der vorliegende Beitrag will diesen gesellschaftlichen Diskussionsprozess unterstützen. Im Folgenden werden zunächst einige herausragende Potentiale verknüpfter, prozessproduzierter Gesundheitsdaten herausgearbeitet, sodann werden Ansatzpunkte für einen angemessenen Regulierungsrahmen skizziert. Dabei richten wir den Blick auch auf den australischen Bundesstaat Westaustralien, wo in den letzten Jahrzehnten eine in vielerlei Hinsicht beispielhafte verknüpfte Dateninfrastruktur aufgebaut wurde.

## 1 Potentiale verknüpfter Gesundheitsdaten

Die Verknüpfung und die zentrale Speicherung von Gesundheitsdaten bieten zunächst einige unmittelbare Vorteile für die Patient\*innen selbst: Es entsteht ein individuelles Gut. Behandlungen können personalisiert und optimiert, Behand-

lungsfehler und Doppelbehandlungen vermieden werden. Genau dieser Prozess soll in Deutschland ab 2021 mit der elektronischen Patient\*innenakte forciert werden.

Wenn es aber, wie hier, um Gesundheitsdaten als öffentliches Gut geht, stehen die Verfügbarkeit und Nutzung verknüpfter digitaler Gesundheitsdaten für große Bevölkerungsteile oder sogar für die Bevölkerung insgesamt im Mittelpunkt. Für die einzelne Patientin ergibt sich aus der Nutzung ihrer Daten oftmals kein nennenswerter unmittelbarer Nutzen; der gesellschaftliche Nutzen kann aber erheblich sein. Darin besteht die zentrale Parallele zu anderen Kollektivgütern, beispielsweise zur „Herdenimmunität“, die durch eine hinreichend hohe Impfquote erreicht wird. Ist eine hinreichend große Zahl an Personen etwa gegen Masern geimpft, bringt die Impfung für den Einzelnen allenfalls überschaubare Vorteile – denn die Gefahr einer Infektion ist aufgrund der Herdenimmunität gering. Die Risiken einer Impfung können ihren Nutzen aus individueller Perspektive sogar überwiegen, da Impfungen immer gewisse – wenn auch vergleichsweise sehr geringe – Nebenwirkungen haben können. Entscheidet man sich aber gegen die Impfung, so gefährdet man das öffentliche Gut der Herdenimmunität, welches ja nur dadurch erreicht werden kann, dass sich eine hinreichend große Zahl von Personen impfen lässt.

Kehren wir zurück zu Gesundheitsdaten als öffentlichem Gut. Worin genau besteht der gesellschaftliche Nutzen? Warum werden diese Daten überhaupt benötigt, schließlich gibt es doch bereits eine Reihe von Befragungen, die mehr oder weniger umfangreiche Informationen zur Gesundheit der Befragten erheben? Beispiele hierfür sind Studien des Robert Koch-Instituts, der *NAKO Gesundheitsstudie* und der *Survey of Health, Ageing and Retirement in Europe* (SHARE). Auch viele weitere sozialwissenschaftliche Surveys wie etwa das *Sozio-oekonomische Panel* (SOEP) oder das *Nationale Bildungspanel* (NEPS) erheben eine Reihe von Gesundheitsindikatoren. Diese Datensätze sind wichtig und unverzichtbar, enthalten sie doch eine Reihe von Informationen zur individuellen Biographie, zum Haushaltskontext oder zum Gesundheitsverhalten, die in Routinedaten oftmals fehlen. Gleichzeitig haben Routinedaten aber einige entscheidende Stärken, durch die die Datengrundlage für Gesundheitsforschung und -berichterstattung entscheidend bereichert wird. Zu den wichtigsten Vorteilen gehören die Größe der Datensätze, ihre schnelle Verfügbarkeit und ihre Detailliertheit etwa im Hinblick auf objektive Gesundheitsinformationen, Diagnosen, Therapien und Medikationen.

Noch weiter steigern lässt sich der Nutzen von prozessproduzierten Gesundheitsdaten durch die *Verknüpfung* mit weiteren Datenquellen, z. B. aus dem Sterberegister, aus anderen Zweigen der Sozialversicherung wie der Arbeitslosenversicherung oder aus den genannten Bevölkerungsumfragen. Im Idealfall



entsteht so ein detailliertes Bild individueller Lebensverläufe, das bereits genannte „Buch des Lebens“, das die Untersuchung des komplexen Zusammenspiels sozialer und individueller Faktoren bei der Entstehung und Bewältigung von Krankheiten und strukturellen gesundheitlichen Ungleichheiten ermöglicht.

Die Größe von Routinedatensätzen erlaubt unter anderem eine deutlich feingliedrigere räumliche Unterteilung als klassische Umfragedaten. Damit lassen sich Fragestellungen bearbeiten, die mit herkömmlichen Daten kaum zu beantworten wären, etwa zum Zusammenhang zwischen Wohnumgebung und Gesundheit (Engemann et al. 2019). Große Fallzahlen sind zudem eine oftmals entscheidende Voraussetzung für die Untersuchung von Kausalzusammenhängen mit modernen quasi-experimentellen Verfahren wie der Methode der Instrumentalvariablen oder der Regressions-Diskontinuitäts-Analyse (Gross et al. 2017; Hofmann & Mühlenweg 2017). Diese Verfahren sind in der Regel auf Vergleiche zwischen sehr spezifischen Subgruppen angewiesen, um daraus allgemeine Schlussfolgerungen hinsichtlich der kausalen Effekte von (gesundheits-) politischen Interventionen oder anderen Faktoren zu ziehen. In regulären Befragungen sind die relevanten Subgruppen oftmals nicht groß genug, um empirische Befunde statistisch hinreichend abzusichern. Ähnliches gilt – ganz unabhängig von den spezifischen Anforderungen quasi-experimenteller Untersuchungsdesigns – für seltene, in ihren Folgen für die Betroffenen oftmals aber verheerende Erkrankungen wie z.B. bestimmte Krebsarten oder post-traumatische Belastungsstörungen. Auch hier ermöglicht die Größe von Routinedatensätzen tiefgehende Analysen zu umwelt- und verhaltensbezogenen Bestimmungsfaktoren, die mit anderen Datenquellen schlichtweg nicht zu realisieren sind.

Eine weitere wichtige Stärke von Routinedaten besteht in der besseren Repräsentation bestimmter sozialer Gruppen und Krankheitsbilder (vgl. hierzu auch Stanley et al. 2011). Zu denken ist hier etwa an Personen mit schweren psychischen oder körperlichen Erkrankungen oder an Opfer häuslicher Gewalt. In herkömmlichen Befragungen sind diese Gruppen deutlich unterrepräsentiert. Zudem ist davon auszugehen, dass Fragen zu derartigen sensiblen Themen oftmals nicht oder nicht wahrheitsgemäß beantwortet werden. Die entsprechenden Potentiale (verknüpfter) Gesundheitsdaten werden zum Beispiel in verschiedenen Arbeiten auf Grundlage der unten näher beschriebenen westaustralischen Dateninfrastruktur deutlich (Hofmann & Mühlenweg 2017). So konnten O'Donnell und Koautor\*innen (O'Donnell et al. 2009; O'Donnell et al. 2010) verschiedene negative Langfristeffekte des mütterlichen Gesundheitsverhaltens und Drogenkonsums während der Schwangerschaft auf das spätere Wohlergehen der Kinder nachweisen.

Ein weiterer wichtiger Vorteil verknüpfter digitaler Routinedaten ist ihre schnelle Verfügbarkeit, eine funktionierende und effiziente Infrastruktur vorausgesetzt. So können sie einen wichtigen Beitrag zum sogenannten „nowcasting“, zur Gesellschaftsbeobachtung in Echtzeit, leisten (Lazer & Radford 2017). In Verbindung mit anderen Datenquellen etwa zum Suchverhalten im Internet, können Routinedaten zum Beispiel effizientere Überwachungs- und Frühwarnsysteme für die Ausbreitung von ansteckenden Krankheiten wie der saisonalen Influenza oder COVID-19 ermöglichen. Mögliche Bedrohungen für die öffentliche Gesundheit können so schneller und zuverlässiger erkannt und geeignete Gegenmaßnahmen ergriffen werden.

Auch zu anderen Formen der Gesundheitsberichterstattung und -überwachung können Routinedaten wichtige Beiträge leisten. Auffällige Häufungen von Krankheiten können auf unentdeckte schädliche Umwelteinflüsse oder unbekannte Nebenwirkungen von Medikamenten hindeuten und mit Routinedaten schneller erkannt werden. Das „Sentinel“-Programm der US-amerikanischen *Food and Drug Administration* (FDA) etwa nutzt Routinedaten, um die Wirksamkeit und mögliche unerwünschte Nebenwirkungen von Medikamenten nach der Zulassung zu überwachen (OECD 2019). Durch die Nutzung digitaler Routinedaten kann diese Überwachung effektiver und kostengünstiger erfolgen und herkömmliche Verfahren, die etwa auf der pro-aktiven Meldung von Komplikationen durch behandelnde Ärzt\*innen basieren, wirksam ergänzen. Das Programm ist zudem ein gutes Beispiel für eine Architektur, die Erkenntnis- und Datenschutzinteressen in Einklang zu bringen versucht. Dank einer einheitlichen, harmonisierten Struktur der einzelnen dezentralen Datenbanken etwa von Krankenhäusern oder Versicherungen ist die Zusammenführung sämtlicher Informationen in einer umfassenden Datenbank gar nicht erforderlich. Anfragen der koordinierenden Behörde können unmittelbar von den dezentralen Institutionen bearbeitet werden, so dass nur die Ergebnisse zurückgemeldet und die vollständigen Datensätze nicht geteilt werden müssen. Auch das unten näher beschriebene *West Australian Data Linkage System* wurde bereits für zahlreiche Untersuchungen dieser Art genutzt. Wichtige Beispiele sind Studien zum Auftreten von Geburtsdefekten nach künstlicher Befruchtung (Hansen et al. 2002), zu den Auswirkungen von Langstreckenflügen auf das Thromboserisiko (Kelman et al. 2003) und zu beruflichen Gesundheitsrisiken, etwa durch die Exposition gegenüber kristallinem Silizium (De Klerk et al. 2002).

## 2 Die Nutzung von Gesundheitsdaten bedarf einer breiten gesellschaftlichen Diskussion

Die Verknüpfung von Gesundheitsdaten ist in vielen anderen Ländern bereits weit fortgeschritten (OECD 2015). Wichtige Beispiele sind Kanada, das Vereinigte Königreich, Dänemark, Schweden, die Schweiz und Australien (siehe unten). Fälle von schwerem Missbrauch oder Datendiebstahl sind bisher kaum bekannt, was aber nicht bedeutet, dass solche Ereignisse unvorstellbar sind – und natürlich können wir auch nicht ausschließen, dass es bereits zu derartigen Vorfällen gekommen ist, ohne dass man davon erfahren hätte. Auch in Deutschland ist die Nutzung von Routinedaten schon seit Längerem in verschiedenen Kontexten möglich. Wichtige Daten der Arbeitslosen- und Rentenversicherung etwa können über die Forschungsdatenzentren des Instituts für Arbeitsmarkt- und Berufsforschung und der Gesetzlichen Rentenversicherung für Forschungszwecke genutzt werden. Auch die Nutzung von Daten der gesetzlichen Krankenkassen ist prinzipiell möglich; der Datenzugang erfolgt jedoch über die einzelnen gesetzlichen Krankenkassen. Das ist nicht nur sehr aufwendig, sondern schmälert den Nutzen der Daten auch deshalb, weil viele Projekte sich auf die Versicherten einer einzelnen Krankenkasse beschränken müssen, was sich nachteilig auf die Repräsentativität der Untersuchungspopulation auswirkt.

Obwohl Routinedaten also bereits genutzt werden, hinkt die gesellschaftliche Diskussion über das Für und Wider der Datennutzung und über angemessene Regeln für den Umgang mit Gesundheitsdaten in Deutschland weit hinterher. Sehr deutlich zeigte sich dies zuletzt in der massiven und durchaus berechtigten Kritik an Teilen des *Digitale-Versorgung-Gesetzes* aus dem Jahr 2019, insbesondere an der Schaffung eines Forschungsdatenzentrums, das die Abrechnungsdaten der verschiedenen gesetzlichen Krankenkassen in pseudoanonymisierter Form zusammenführen und für die Forschung verfügbar machen soll.

Deutschland braucht in dieser Frage den öffentlichen Diskurs. Wird er nicht geführt, bleiben die Potentiale verknüpfter Gesundheitsdaten weiter weitgehend ungenutzt, zum Nachteil von Forschung, Gesundheitssystem und Patient\*innen. Trotz – oder gerade wegen – der Chancen, die die Nutzung von verknüpften digitalen Gesundheitsdaten bietet, sollten aber auch die erheblichen Risiken nüchtern und vorbehaltlos anerkannt werden. Ohne Frage können die unsachgemäße Nutzung oder gar Veröffentlichung sensibler Gesundheitsdaten gravierende Folgen haben, etwa in Form von sozialer Stigmatisierung oder Diskriminierung auf dem Arbeits- und Versicherungsmarkt. Datensicherheit und Datenschutz müssen daher höchste Priorität haben.

Wichtig ist, dass dieser gesellschaftliche Diskurs nicht bei Null anfangen, dass das Rad nicht neu erfunden werden muss. Die Erfahrungen anderer Länder und Expert\*innenempfehlungen zur Regulierung von Gesundheitsdaten, wie sie in den letzten Jahren etwa im Rahmen der Gesellschaft für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) erarbeitet wurden, können wichtige Impulse für die Diskussion in Deutschland liefern. So listet die offizielle Empfehlung des OECD-Rates zur Regulierung von Gesundheitsdaten aus dem Jahr 2016 zwölf Kernprinzipien für die nationalen Regulierungsrahmen auf (OECD 2019). Beteiligung, Transparenz und regelmäßige Überprüfungen unter Einbindung verschiedener Interessengruppen sind zentrale Elemente dieser Rahmenvorschläge. Grundsätzlich sollte Dritten, auch gemeinnützigen wissenschaftlichen Einrichtungen, kein genereller Zugang zu den Gesundheitsdaten gewährt werden. Der Zugang sollte stets zweckgebunden und nur nach hinreichender Darlegung des gesellschaftlichen Nutzens erfolgen. Ausreichende Infrastrukturen und Expertise im Bereich Datensicherheit sollten notwendige Voraussetzungen sein, damit eine Institution und ihre Mitarbeiter\*innen Zugang zu diesen sensiblen Daten erhalten können. Alternativ wäre denkbar, dass der Zugang nur an ausgewählten Standorten mit besonderer Sicherheitsarchitektur erfolgen kann, so wie dies bereits jetzt der Regelfall bei der Arbeit mit Routinedaten aus anderen Zweigen der Sozialversicherung ist.

Um zu veranschaulichen, wie ein zukunftsfähiges Konzept für eine verbesserte Datenbasis aussehen und die adäquate Berücksichtigung verschiedener gesellschaftlicher Interessen gewährleistet werden könnte, stellen wir im Folgenden das westaustralische System einer verknüpften Dateninfrastruktur vor.

### 3 Westaustralien als Best-Practice-Beispiel für eine moderne verknüpfte Dateninfrastruktur

Der australische Bundesstaat Westaustralien hat im Laufe der letzten Jahrzehnte eine in vielerlei Hinsicht beispielhafte vernetzte Dateninfrastruktur geschaffen (Holman et al. 2008). Das *Western Australia Data Linkage System* (WADLS) verknüpft Gesundheits- und Sozialdaten aus 39 Quellen. Neben Gesundheitsdaten (u. a. aus dem Krankenhaussystem und dem Sterberegister) beinhaltet das WADLS eine Vielzahl weiterer Routinedaten etwa aus den Bereichen Bildung, Justiz- und Strafvollzug, Kinderschutz und Wohnen. Auch wenn die einzelnen Datensätze in Deutschland nur wenigen geläufig sein dürften, verdeutlicht Abbildung 1a sehr gut, wie breit das Spektrum der im Rahmen des WADLS verknüpfbaren Daten inzwischen ist.

Im August 2019 umfassten die Daten mehr als 117 Millionen Einträge, die sich auf insgesamt fast 4,5 Millionen „Ketten“ („chains“) verteilen und einen Zeitraum von mehr als 50 Jahren abdecken. Im Regelfall ist eine „Kette“ von verknüpfbaren Einträgen genau einem Individuum zuzuordnen und jedem Individuum genau eine Kette. Grundsätzlich ist die gesamte westaustralische Bevölkerung von derzeit etwa 2,8 Millionen Menschen im WALDS erfasst (Government of Western Australia – Department of Health 2020b). Für Geburten ab dem Jahr 1975 ist zudem eine Verknüpfung von Eltern und Kindern möglich, so dass die Daten optimale Voraussetzungen bieten, um die Weitergabe gesundheitlicher und sozialer Ungleichheiten von einer Generation zur nächsten zu untersuchen. Besonders interessant – aber auch datenschutztechnisch sensibel – ist in diesem Zusammenhang die zunehmende Ergänzung der Daten mit genetischen Informationen.

Neben den verschiedenen Routinedatensätzen wurden inzwischen eine Reihe von Bevölkerungsumfragen, Kohortenstudien, Fall-Kontroll-Studien und randomisierten kontrollierten Studien in das WADLS integriert. Die Verknüpfung mit den Routinedaten kann die Potentiale solcher sozialwissenschaftlicher und medizinischer Datensätze deutlich erhöhen: So können Querschnittbefragungen mit Hilfe der Routinedaten mit längsschnittlichen Lebensverlaufsinformationen angereichert werden, und Personen, die vorzeitig aus Längsschnittbefragungen ausscheiden, können auf Grundlage der Routinedaten zumindest eingeschränkt weiterverfolgt werden (Stanley et al. 2011).

Der Erfolg dieses herausragenden Dateninfrastrukturprojekts zeigt sich darin, dass seit dem Jahr 2009 am Aufbau einer vergleichbaren Infrastruktur für das gesamte Land gearbeitet wird, auch wenn sich dieser Prozess auf Grund verschiedener technischer und administrativer Hürden als recht langwierig herausgestellt hat.

Zentrale Voraussetzung für den Erfolg des WADLS war und ist die frühzeitige und konsequente Einbeziehung verschiedenster gesellschaftlicher Gruppen. Die Zusammenarbeit von Forschung, Gesundheitswesen, Sozialversicherung, Verwaltung und Patient\*innenvertretungen war entscheidend für die Entwicklung einer effizienten technischen Infrastruktur und für die Etablierung von Sicherheits- und Nutzungsprotokollen mit breiter gesellschaftlicher Akzeptanz (Holman et al. 2008). Ein wichtiger Baustein der dauerhaften Patient\*innenbeteiligung ist das Amt einer/eines vollzeitbeschäftigten Patient\*innenbeauftragten für die gegenseitige Vermittlung zwischen Forschung und Öffentlichkeit in Fragen des Datenschutzes und der Datennutzung. Die Sichtbarkeit von Forschungsergebnissen auf Grundlage des WADLS, unter anderem durch eine wöchentliche Beilage in der wichtigsten Tageszeitung des Bundesstaats, unterstreicht den Kollektivgutcharakter des Systems und fördert so seine Akzeptanz.

## Welche Daten sind verknüpft?

### Kerndatensätze

- Hospital Morbidity Data Collection
- Mental Health Information System
- Emergency Department Data Collection
- WA Cancer Registry
- Midwives Notifications System
- Death Registrations
- Birth Registrations
- Marriage Registrations (nur zu Verknüpfungszwecken; nicht zur Veröffentlichung verfügbar)
- WA Electoral Roll

### WA Gesundheit

- Home and Community Care
- Aged Care Assessment Program
- WA Notifiable and Infectious Diseases Database
- Monitoring of Drugs of Dependence
- State Trauma Registries
- WA Register of Developmental Anomalies
- iPharmacy and ASCRibe hospital dispensary data
- Child Development Information System
- BreastScreen WA
- Reproductive Technology Register
- Elective Surgery Waitlist
- Cervical Screening Register

### WA Regierung

- Department of Child Protection
- Department of Education
- School Curriculum and Standards Authority
- Department of Corrective Services
- Disability Services Commission
- Department of Housing
- Department of the Attorney General
- Department of Transport

### Andere Organisationen

- Silver Chain
- Insurance Commission
- Main Roads WA
- Intellectual Disability
- St John Ambulance
- Brightwater Care Group

## Das Trennungsprinzip

Die in einem elektronischen Datensatz enthaltenen Informationen können in drei Kategorien unterteilt werden:

**Informationen zur Identifizierung - Wer ist die Person?**  
z.B. Name, vollständiges Geburtsdatum, Adresse

**Versorgungsdaten - Was ist mit dieser Person geschehen?**  
z.B. Diagnose, Details zur Behandlung, Testergebnisse

**Weitere Informationen, die keine unmittelbare Identifikation einzelner Personen ermöglichen**  
z.B. Datum der (Gesundheits-)Dienstleistung, Geschlecht, Postleitzahl

### Wer kann die Daten einsehen?

**Informationen zur Identifizierung:**  
Diese werden von den Verknüpfungsbeauftragten der DLB genutzt, um verschiedene Datensätze über bleibende WADLS-Identifikationsnummern („linkage keys“) zu verknüpfen. Die Endnutzer\*innen der verknüpften Daten können diese Nummern niemals einsehen.

**Nicht-identifizierende (Versorgungs-) Informationen**  
Diese Informationen werden von den Endnutzer\*innen mit Hilfe projektspezifischer Verknüpfungsnummern zusammengeführt und ausgewertet. Die Verknüpfungsbeauftragten der DLB können diese Informationen nicht einsehen.

**Informationen, die keine unmittelbare Identifikation einzelner Personen ermöglichen:**  
Diese Informationen können sowohl für den Verknüpfungsprozess als auch für die Datenauswertung nötig sein. Sie können daher von den Verknüpfungsbeauftragten der DLB, sofern sie für das jeweilige Forschungsprojekt nötig sind, aber auch von den Endnutzer\*innen eingesehen werden.

Durch die Einhaltung des Trennungsprinzips können die DLB und die Endnutzer\*innen der verknüpften Datensätze den maximalen Nutzen aus diesen wertvollen Informationsquellen ziehen und gleichzeitig die Privatsphäre der Personen schützen.

**Abb. 1a** (obere Hälfte): Verknüpfte Datenquellen im WADLS.

**Abb. 1b** (untere Hälfte): Das Trennungsprinzip des WADLS.

Quelle: Government of Western Australia – Department of Health 2020a. Übersetzung durch die Autor\*innen. WA = Western Australia/Westaustralien. DLB = *Data Linkage Branch*, die für die Datenverknüpfung zuständige Abteilung des westaustralischen Gesundheitsministeriums.



## 4 Datenschutz im WADLS

Die Datenschutzstrategie des WADS beginnt mit dem allgemeinen nationalen Regulierungsrahmen und der Überwachung durch die nationale Datenschutzbeauftragte. Verknüpfung, Speicherung und Auswertung der Daten unterliegen klaren und strengen Regeln. Forschungsprojekte, die das WADLS nutzen, werden in der Regel mehrfach geprüft, unter anderem von den Ethikkommissionen der beteiligten Forschungseinrichtungen und des Gesundheitsministeriums sowie von den Datenschutzbeauftragten der datenerhebenden Institutionen. Personen, die im Rahmen des Verknüpfungsprozesses Zugriff auf besonders vertrauliche Informationen erhalten, müssen hohe Anforderungen erfüllen (z. B. Überprüfung möglicher Vorstrafen).

Ein zentrales Instrument zum Schutz der Privatsphäre ist das in Abbildung 1b dargestellte sogenannte „Trennungsprinzip“ („separation principle“; vgl. dazu auch die ausführliche Darstellung in Kelman et al. 2002). Der Kern dieses Prinzips besteht darin, dass die einzelnen prinzipiell verknüpfbaren Datensätze getrennt voneinander gespeichert und verwaltet werden. Im WADLS existiert somit kein vollständiger integrierter Datensatz, der sämtliche Informationen aus allen prinzipiell verfügbaren Datensätzen enthielte. Die Datenverknüpfung erfolgt anlassbezogen (z. B. für ein bestimmtes Forschungsprojekt) und jeder einzelne Verknüpfungsvorgang beschränkt sich auf die für den jeweiligen Anlass notwendigen Datensätze und Informationen. Besonders wichtig ist in diesem Zusammenhang, dass alle Informationen, die eine unmittelbare Identifikation von Einzelpersonen ermöglichen (z. B. Adresse, Geburtsdatum Sozialversicherungsnummer), grundsätzlich getrennt von den individuellen Gesundheits- und Sozialdaten gespeichert werden. Die Verknüpfung der verschiedenen Datenquellen erfolgt über generische WADLS-Identifikationsnummern („linkage keys“) ohne Bedeutung außerhalb des Systems (und somit z. B. nicht über Kennziffern, die im Rahmen der Sozialversicherung verwendet werden). Informationen, die eine eindeutige Identifikation von Personen erlauben (z. B. Sozialversicherungsnummer, Name, Geburtsdatum und Wohnort) werden nur für die anfängliche Erstellung der generischen Identifikationsnummer verwendet und danach von dieser getrennt.

Im Einzelnen umfasst das Datenverknüpfungsprotokoll im WADLS damit die folgenden zentralen Elemente und Schritte:

1. Ausgewählte Verknüpfungsbeauftragte („linkage officers“) erstellen und pflegen eine Verknüpfungsdatenbank mit eindeutigen bleibenden WADLS-Identifikationsnummern, die in jedem zum System gehörenden Datensatz vorhanden sind und so die Verknüpfung ermöglichen. Informationen, die

eine eindeutige Identifikation von Personen erlauben, werden nur bei der Erstellung dieser WADLS-Identifikationsnummern verwendet und anschließend nicht mehr zusammen mit diesen gespeichert.

2. Endnutzer\*innen (z.B. Forscher\*innen) beantragen die Verknüpfung bestimmter Informationen aus den einzelnen am System beteiligten Datensätzen für spezifische Anlässe und Projekte. Vollständige Datensätze werden grundsätzlich nicht an die Endnutzer\*innen ausgeliefert.
3. Die Verknüpfungsbeauftragten erstellen eine projektspezifische Zuordnung von bleibenden WADLS-Identifikationsnummern zu anlassbezogenen (projektspezifischen) Identifikationsnummern und übermitteln diese an die Verwalter\*innen der einzelnen Datensätze.
4. Die Verwalter\*innen erstellen einen Auszug aus dem verwalteten Datensatz, der nur die für das jeweilige Projekt nötigen Informationen enthält (siehe Schritt 2). Die unveränderlichen WADLS-Identifikationsnummern werden von den Verwalter\*innen gemäß der Zuordnung aus Schritt 3 in die projektspezifischen Identifikationsnummern überführt. Die unveränderlichen WADLS-Identifikationsnummern werden vor der Weitergabe aus den Daten entfernt.
5. Die verschiedenen Teildatensätze aus Schritt 4 werden von den Endnutzer\*innen unter Verwendung der anlassbezogenen Verknüpfungsnummern zusammengeführt. Die Verwalter\*innen der einzelnen Datensätze sind somit nicht in der Lage, diese mit Informationen aus anderen Datensätzen zu verknüpfen. Gleichzeitig wird das Risiko einer Verknüpfung von Personen über verschiedene Projekte hinweg dadurch minimiert, dass den Endnutzer\*innen nur die anlassbezogenen und von Projekt zu Projekt wechselnden Identifikationsnummern vorliegen.

## 5 Was wir gelernt haben – und gemeinsam umsetzen sollten

Das westaustralische Beispiel zeigt, wie eine umfassende nutzer\*innenfreundliche und datenschutzkonforme Infrastruktur für verknüpfte Gesundheits- und Sozialdaten aussehen könnte. Es zeigt zudem, wie sich eine breite gesellschaftliche Akzeptanz einer solchen Infrastruktur erreichen lässt: durch die konsequente Aufklärung über den Nutzen einer gemeinwohlorientierten Verwendung von Gesundheitsdaten und durch die Einbindung verschiedener Interessengruppen in die Entwicklung und Überwachung des Systems. Allerdings verdeutlicht das Beispiel Westaustraliens auch, wie weit wir in Deutschland von einer solchen Infrastruktur entfernt sind und wie viel Aufklärungsarbeit und gesellschaftliche Diskussion noch nötig sind. Deutschland muss den westaustralischen



Ansatz nicht eins zu eins kopieren, aber es wäre fahrlässig, das Thema einer gemeinwohlorientierten Nutzung von prozessproduzierten Gesundheitsdaten nicht endlich entschiedener anzugehen. Die Corona-Pandemie hat uns zuletzt besonders deutlich vor Augen geführt, wie wichtig eine solide Datenlage und darauf basierende wissenschaftliche Erkenntnisse auch für weitreichende politische Entscheidungen sein können.

Die technischen und administrativen Grundvoraussetzungen für den Aufbau einer modernen Dateninfrastruktur sind in Deutschland gegeben: Wichtige Gesundheitsdaten werden routinemäßig digital erfasst und es gibt eine Reihe weiterer Routine- und Umfragesätze, die mit diesen gewinnbringend verknüpft werden könnten. Womöglich wäre die Akzeptanz für einen Ausbau der gemeinwohlorientierten Nutzung von Gesundheitsdaten auch größer als vielfach angenommen. Das Beispiel der Corona-Warn-App, die bis zum 15. April 2021 ca. 27 Millionen Mal heruntergeladen wurde, könnte zumindest ein Indiz dafür sein, dass weite Teile der Bevölkerung mit der Nutzung ihrer Daten einverstanden sind, wenn diese nach transparenten Regeln erfolgt und ihr ein klar erkennbarer gesellschaftlicher Nutzen gegenübersteht.

## Literatur

- De Klerk, Nicholas H./Ambrosini, Gina L./Musk, A. William (2002): „Crystalline Silica Exposure and Major Health Effects in Western Australian Gold Miners“. In: *The Annals of Occupational Hygiene* 46(Supplement 1), S.1–3.
- Dunn, Halbert L. (1946): „Record Linkage“. In: *American Journal of Public Health and the Nation's Health* 36(12), S. 1412–1416.
- Engemann, Kristine/Pedersen, Carsten B./Arge, Lars et al. (2019): „Residential Green Space in Childhood Is Associated with Lower Risk of Psychiatric Disorders from Adolescence into Adulthood“. In: *Proceedings of the National Academy of Sciences of the USA* 116(11), S. 5188–5193.
- Government of Western Australia – Department of Health (2020a): „Data Linkage Branch Material“. <https://www.datalinkage-wa.org.au/resources/data-linkage-branch-material/>, besucht am 28. 8. 2020.
- Government of Western Australia – Department of Health (2020b): „The WA Data Linkage System“. <https://www.datalinkage-wa.org.au/data/wa-data-linkage-system>, besucht am 28. 8. 2020.
- Gross, Christiane/Hofmann, Sarah/Mühlenweg, Andrea et al. (2017): „Theoretische und empirische Perspektiven auf Bildung, Gesundheit und Arbeitsfähigkeit – ein interdisziplinärer Überblick“. In: *Sozialer Fortschritt* 66(1), S. 3–30.
- Hansen, Michèle/Kurinczuk, Jennifer J./Bower, Carol et al. (2002): „The Risk of Major Birth Defects After Intracytoplasmic Sperm Injection and in vitro Fertilization“. In: *New England Journal of Medicine* 346(10), S. 725–730.

- Hofmann, Sarah/Mühlenweg, Andrea (2017): „Eine Perspektive für Analysen des Zusammenhangs von Bildung und Gesundheit: Potential und Limitationen von Routinedaten der Gesetzlichen Krankenversicherung (GKV)“. In: *Sozialer Fortschritt* 66(1), S. 49–75.
- Holman, C D'Arcy J./Bass, John A./Rosman, Diana L. et al. (2008): „A Decade of Data Linkage in Western Australia: Strategic Design, Applications and Benefits of the WA Data Linkage System“. In: *Australian Health Review* 32(4), S. 766–777.
- Kelman, Christopher W./Bass, Alfred J./Holman, C D'Arcy J. (2002): „Research Use of Linked Health Data – A Best Practice Protocol“. In: *Australian and New Zealand Journal of Public Health* 26(3), S. 251–255.
- Kelman, Christopher W./Kortt, Michael A./Becker, N. G. et al. (2003): „Deep Vein Thrombosis and Air Travel: Record Linkage Study“. In: *BMJ* 327, 1072.
- Lazer, David/Radford, Jason (2017): „Data ex Machina: Introduction to Big Data“. In: *Annual Review of Sociology* 43(1), S. 19–39.
- O'Donnell, Melissa/Nassar, Natasha/Leonard, Helen et al. (2009): „Increasing Prevalence of Neonatal Withdrawal Syndrome: Population Study of Maternal Factors and Child Protection Involvement“. In: *Pediatrics* 123(4), S. e614–e621.
- O'Donnell, Melissa/Nassar, Natasha/Leonard, Helen et al. (2010): „Characteristics of Non-Aboriginal and Aboriginal Children and Families with Substantiated Child Maltreatment: A Population-based Study“. In: *International Journal of Epidemiology* 39(3), S. 921–928.
- OECD (2015): *Health Data Governance: Privacy, Monitoring and Research – Policy Brief*. Paris: Organisation for Economic Co-operation and Development, <https://www.oecd.org/health/health-systems/Health-Data-Governance-Policy-Brief.pdf>, besucht am 11.11.2020.
- OECD (2019): *Recommendation of the Council on Health Data Governance*. Paris: Organisation for Economic Co-operation and Development, <https://www.oecd.org/health/health-systems/Recommendation-of-OECD-Council-on-Health-Data-Governance-Booklet.pdf>, besucht am 11.11.2020.
- Stanley, Fiona/Glauert, Rebecca/McKenzie, Anne et al. (2011): „Can Joined-up Data Lead to Joined-up Thinking? The Western Australian Developmental Pathways Project“. In: *Healthcare Policy* 6(Spec. Issue), S. 63–73.