REDEKER | SELLNER | DAHS

Bericht zur Datenschutz-Folgenabschätzung

- -ENTWURF-

Datum:	Entwurf vom 19.08.2022
Betreff:	Datenschutz-Folgenabschätzung für die Datenverarbeitungsvorgänge, die in Umsetzung der §§ 303a bis 303f SGB V zur Datentransparenz im Forschungsdatenzentrum (FDZ) im Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) erfolgen sollen
Reg.-Nr.:	76 20 1304
Verfasser:	[redacted]

[redacted]

Inhaltsverzeichnis

[redacted]

I. Vorbemerkungen

Die nachstehenden Ausführungen enthalten den Bericht zur durchgeführten Datenschutz-Folgenabschätzung (DSFA) für die Datenverarbeitungsvorgänge, die im Forschungsdatenzentrum (FDZ) im Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) in datenschutzrechtlicher Verantwortlichkeit des BfArM erfolgen. Anknüpfungspunkt hierfür sind die Neuregelungen zur sog. Datentransparenz im SGB V sowie in der Datentransparenzverordnung (DaTraV).

Teilweise sind für die Verarbeitung personenbezogener Daten und deren Risikobeurteilung wichtige Abläufe und Systeme noch nicht abschließend definiert. Dies kann nur sukzessive im Zuge von mehreren Ausbaustufen erfolgen, deren Abläufe frühestens im Laufe des Jahres 2023 definiert sein werden. Insofern wurde die DSFA vor dem Hintergrund der feststehenden Systemfunktionalitäten durchgeführt; im Übrigen wurde, um auch die Abläufe und Systeme der zukünftigen Ausbaustufen bereits jetzt weitestgehend zu berücksichtigen, sowohl im Rahmen der Beschreibung der Verarbeitungsvorgänge (Art. 35 Abs. 7 lit. a DS-GVO) als auch bei der Risikoanalyse (Art. 35 Abs. 7 lit. c DS-GVO) von gewissen Prämissen ausgegangen.

Eine erneute Durchführung der DSFA ist stets dann erforderlich, wenn sich für die Verarbeitung personenbezogener Daten relevante Verfahren und Systeme ändern oder – insbesondere mit Blick auf die zukünftigen Ausbaustufen – neue eingeführt werden. Dies kann regelmäßig zu neuen Risiken und einer entsprechend anderen Risikobewertung führen. Eine Aktualisierung der DSFA ist ferner erforderlich, wenn sich die rechtlichen Rahmenbedingungen ändern. Insofern handelt es sich bei diesem DSFA-Bericht um einen Entwurf, der angesichts des iterativen Umsetzungsprozesses im FDZ regelmäßig angepasst werden muss.

II. Datenschutzrechtliche Verantwortlichkeit gemäß Art. 4 Nr. 7 DS-GVO

Gemäß Art. 4 Nr. 7 DS-GVO ist datenschutzrechtlich Verantwortlicher die natürliche oder juristische Person, Behörde, Einrichtung oder andere Stelle, die allein oder gemeinsam mit anderen über die Zwecke und Mittel der Verarbeitung von personenbezogenen Daten entscheidet. Dies ist hinsichtlich der im FDZ erfolgenden Datenverarbeitungsvorgänge vorliegend das BfArM als zuständige Behörde. Das BfArM hat die Bundesdruckerei GmbH (BDR) damit beauftragt, die Infrastruktur des FDZ-Systems zu entwickeln und – jedenfalls im Rahmen der ersten Ausbaustufe (sog. Pilotbetrieb) – zu betreiben. Insoweit setzt das BfArM die BDR als Auftragsverarbeiter ein und hat mit ihr einen entsprechenden Auftragsverarbeitungsvertrag gemäß Art. 28 Abs. 3 DSGVO abgeschlossen. Welche Dienstleister das BfArM im Rahmen der nächsten Ausbaustufen des Betriebs des FDZ als Auftragsverarbeiter einsetzt, ist noch nicht abschließend festgelegt.

Datenschutzrechtlich verantwortliche Stelle:	Bundesinstitut für Arzneimittel und Medizinprodukte
Anschrift:	Kurt-Georg-Kiesinger-Allee 3 53175 Bonn
Telefon:	+49 (0)228-99-307-[redacted]
E-Mail-Adresse:	[redacted]
Leiter/Ansprechpartner:	Dr. Steffen Heß
Datenschutzbeauftragter:	[redacted]

Abzugrenzen hiervon sind Verarbeitungsvorgänge unter datenschutzrechtlicher Verantwortung einzelner Krankenkassen (1), unter der Verantwortung des Spitzenverbandes Bund der Krankenkasse (2) und des Robert-Koch-Instituts als Vertrauensstelle (3). Mögen auch entsprechende Datenverarbeitungsvorgänge – insbesondere unter der datenschutzrechtlichen Verantwortlichkeit der Vertrauensstelle des RKI – Auswirkungen auf die hiesige Risikoanalyse haben, trifft die gemäß Art. 35 Abs. 1 DS-GVO formulierte Pflicht zur Durchführung einer DSFA jeden o.g. Verantwortlichen getrennt.

III. Beschreibung der geplanten Verarbeitungsvorgänge und der Zwecke der Verarbeitung, Art. 35 Abs. 7 lit. a DS-GVO

1. Vorbemerkungen

Im SGB V regeln die §§ 303a bis 303f die sog. Datentransparenz. Damit sollen Rahmenbedingungen geschaffen werden, um umfangreiche (Sozial-) Datensätze für die Forschung zur Verbesserung der Gesundheitsversorgung zugänglich zu machen. Hierzu können Sozialdaten –

überwiegend in aggregierter und anonymisierter Form – bestimmten, in § 303e Abs. 1 SGB V enumerativ aufgelisteten Institutionen („Nutzungsberechtigte“) unter engen Voraussetzungen zur Verfügung gestellt werden.

Das Verfahren der Datentransparenz wurde mit Inkrafttreten des Digitale-Versorgung-Gesetz vom 9. Dezember 2019 und der DaTraV vom 19. Juni 2020 neu geregelt. Hierdurch wird im Vergleich zur bisherigen Rechtslage insbesondere das Datenangebot erweitert und dessen Aktualität gesteigert. Vor Inkrafttreten der Neuregelungen wurden zur Herstellung der Datentransparenz lediglich die Daten verarbeitet, die von den Krankenkassen im Zuge des morbiditätsorientierten Risikostrukturausgleichs (Morbi-RSA) erhoben wurden. Die Vorschriften zur Datentransparenz – namentlich § 303b Abs. 1 SGB V a.F. – verwiesen nach alter Rechtslage insoweit auf den Katalog des § 268 Abs. 3 S. 14 i.V.m. S. 1 Nr. 1 bis 7 SGB V in der bis zum 31.03.2020 gültigen Fassung.

Unter Geltung der neuen Vorschriften wird der Datensatz zur Herstellung von Datentransparenz nunmehr von der Datengrundlage für die Durchführung des Morbi-RSA getrennt. Gegenüber der alten Rechtslage unverändert geblieben ist indes, dass es sich bei zur Herstellung von Datentransparenz verarbeiteten Sozialdaten um Daten im Sinne von § 67 Abs. 2 SGB X handelt, die den gesetzlichen Krankenkassen *ohnehin vorliegen*. Die gesetzlichen Krankenkassen erheben sie ursprünglich bei den Versicherten zum Zwecke des Vertragsschlusses oder bei den Leistungserbringern zum Zwecke der Abrechnung. Diese Sozialdaten werden, soweit sie im FDZ zur Herstellung von Datentransparenz verarbeitet werden, als DaTraV-Daten bezeichnet.

Neben den gesetzlichen Krankenkassen sind die weiteren, am Verfahren zur Herstellung von Datentransparenz beteiligten Akteure der *Spitzenverband Bund der Krankenkassen* als Datensammelstelle, eine im Robert-Koch-Institut (RKI) zu errichtende *Vertrauensstelle*, das *FDZ*, das im Rahmen dieser DSFA im Fokus steht, und die *Nutzungsberechtigten*. Verarbeitungen personenbezogener Daten, für welche allein die anderen Akteure datenschutzrechtlich verantwortlich sind, blieben im Rahmen der DSFA weitgehend außer Betracht, soweit sie sich nicht auch auf die beim FDZ erfolgenden Datenverarbeitungsvorgänge auswirken.

Die Aufgaben des FDZ führt das *BfArM* aus. Es wird räumlich, organisatorisch und personell eigenständig geführt. Es unterliegt dem Sozialgeheimnis nach § 35 SGB I und untersteht der Rechtsaufsicht des Bundesministeriums für Gesundheit.

2. Maßgebliche Zwecke der Datenverarbeitungen

Die Verarbeitungen personenbezogener Daten, die im Zusammenhang mit den Vorschriften zur Datentransparenz vorgenommen werden, dienen der zielgerichteten Forschung im Gesundheitsbereich und Verbesserung der medizinischen Versorgung.[1] Auf diese Weise werden neue Erkenntnisse über Erkrankungen und Behandlungsmethoden erzielt, die sich positiv auf die Sterblichkeit und das Wohlbefinden der Menschen auswirken.[2] Eine Stärkung der Versorgungsforschung kann dazu beitragen, dass Krankheiten schneller erkannt und gezielter behandelt werden können. Dies wirkt sich wiederum positiv auf die Gesundheit und den Ernährungszustand der Bevölkerung aus. Durch die Stärkung der Versorgungsforschung wird der Weg bereitet für neue Innovationen, sodass das Gesundheitswesen auf dieser Grundlage optimiert werden kann. Dies trägt dazu bei, dass neue Erkenntnisse in der gesamten Solidargemeinschaft ankommen.

Überdies sind in § 303e Abs. 2 SGB V Zwecke normiert, zu denen die Verarbeitungen personenbezogener Daten erfolgen dürfen. Gemäß § 303e Abs. 2 SGB V dürfen die Nutzungsberechtigten Daten für die folgenden Zwecke verarbeiten:

- Wahrnehmung von Steuerungsaufgaben durch die Kollektivvertragspartner,
- Verbesserung der Qualität der Versorgung,
- Planung von Leistungsressourcen, z.B. Krankenhausplanung,
- Forschung, insbesondere für Längsschnittanalysen über längere Zeiträume, Analysen von Behandlungsabläufen oder Analysen des Versorgungsgeschehens (insbesondere für die Bewertung neuer Behandlungsverfahren im Rahmen einer Beobachtungsstudie oder bei dem Nachweis der positiven Versorgungseffekte digitaler Gesundheitsanwendungen),
- Unterstützung politischer Entscheidungsprozesse zur Weiterentwicklung der gesetzlichen Krankenversicherung,
- Analyse und Entwicklung von sektorenübergreifenden Versorgungsformen sowie von Einzelverträgen der Krankenkassen und

---

1 Drucksache des Deutschen Bundestages 19/13438 vom 23.09.2019, Entwurf eines Gesetzes für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz), S. 34.

2 Referentenentwurf des Bundesministeriums der Gesundheit zur „Verordnung zur Neufassung der Datentransparenzverordnung und zur Änderung der Datentransparenz-Gebührenverordnung“ vom 16.06.2020, S. 23 f.

- Wahrnehmung von Aufgaben der Gesundheitsberichterstattung.

Die Norm bezieht sich expressis verbis lediglich auf die Datenverarbeitungen durch *Nutzungsberechtigte* und nicht durch die anderen Akteure. § 303b Abs. 1 SGB V, der die Krankenkassen adressiert, verweist ausdrücklich auf § 303e Abs. 2 SGB V und bestimmt, dass die Krankenkassen die Daten für die darin genannten Zwecke an den Spitzenverband Bund der Krankenkassen übermitteln dürfen. Weiterhin ordnet § 303d Abs. 1 Nr. 1 SGB V mit Blick auf das FDZ an, dass es die Daten für die Auswertung für Zwecke nach § 303e Abs. 2 SGB V aufbereiten darf. Daraus folgt, dass letztlich alle bei den unterschiedlichen Akteuren stattfindenden Datenverarbeitungen auch zu den in § 303e Abs. 2 SGB festgelegten Zwecken erfolgen, da sie eine notwendige Bedingung für die Bereitstellung der Daten an die Nutzungsberechtigten sind.

Jenseits der Verarbeitung der DaTraV-Daten baut das FDZ gemäß § 303d Abs. 1 Nr. 6 i.V.m. § 9 DaTraV ein öffentliches Antragsregister auf, das (teilweise personenbezogene) Daten zu den antragstellenden Nutzungsberechtigten und zu den Vorhaben, für die die Sozialdaten beim FDZ beantragt wurden, enthält [s. dazu unten A.III.3.c)bb)(2)]. Dies hat den Zweck, die Transparenz der Antragstellungen und -bearbeitungen zu steigern und auf diese Weise weitere potenzielle forschende Nutzungsberechtigte auf das Datenangebot aufmerksam zu machen.

3. Relevante Verarbeitungsvorgänge

a) Arten verarbeiteter personenbezogener Daten

Das FDZ verarbeitet einerseits aus Forschungsgesichtspunkten die DaTraV-Daten, wobei es sich um personenbezogene Daten der Versicherten und der Leistungserbringer handelt. Sie werden im FDZ ausschließlich in pseudonymisierter Form verarbeitet. Welche Daten dies konkret sind, wird nachfolgend im Einzelnen aufgeführt. Andererseits verarbeitet es für die Antragstellung, Bescheidung und Offenlegung Daten der Nutzungsberechtigten, die teilweise auch Personenbezug aufweisen [siehe dazu c) bb)].

Der Fokus der DSFA liegt auf der Verarbeitung der DaTraV-Daten, die Bezug zu den gesetzlich Versicherten als betroffene Personen aufweisen.

Die Verarbeitung von Beschäftigtendaten und sonstigen Daten Dritter, die nicht originär im SGB V und der DaTraV geregelt werden, sind nicht Gegenstand der nachstehenden Ausführungen.

Die vom FDZ verarbeiteten DaTraV-Daten werden von den gesetzlichen Krankenkassen zunächst an den Spitzenverband Bund der Krankenkassen und sodann an das FDZ übermittelt (§ 303b Abs. 1 Satz 1 SGB V). Im Einzelnen handelt es sich dabei gemäß § 303b Abs. 1 Satz 1 SGB V, § 3 Abs. 1 DaTraV um die folgenden Daten:

- das Geburtsjahr, das Geschlecht und die Postleitzahl des Wohnortes des Versicherten,

- Angaben zum Versicherungsverhältnis

  - die Betriebsnummer der Krankenkasse des Versicherten,

  - die Versichertentage je Quartal,

  - die Anzahl der Versichertentage, an denen die versicherte Person ihren Wohnsitz oder gewöhnlichen Aufenthalt außerhalb des Gebietes der Bundesrepublik Deutschland hatte,

  - den Versichertenstatus gemäß den Schlüsselnummern der Statistik über die Versicherten in der gesetzlichen Krankenversicherung KM 1 nach der Allgemeinen Verwaltungsvorschrift über die Statistik in der gesetzlichen Krankenversicherung in Verbindung mit § 79 des Vierten Buches Sozialgesetzbuch,

  - die Anzahl der Versichertentage, an denen die versicherte Person Krankengeld nach den §§ 44 und 45 des SGB V in Anspruch genommen hat,

  - die Anzahl der Versichertentage, an denen die versicherte Person an einem strukturierten Behandlungsprogramm nach § 137f Absatz 1 SGB V teilgenommen hat, einschließlich des jeweiligen Programmkennzeichens des strukturierten Behandlungsprogramms,

  - die Anzahl der Versichertentage, an denen die versicherte Person für den Bereich der ärztlichen Versorgung anstelle von Sach- oder Dienstleistungen nach § 13 Absatz 2 SGB V Kostenerstattung gewählt hat oder Wahltarifen nach § 53 Absatz 4 SGB V in Anspruch genommen hat,

- Kosten- und Leistungsdaten

  - nach den §§ 295 und 295a SGB V zur ambulanten Versorgung: Betriebsstättennummer, lebenslange Arzt- oder Zahnarztnummer gemäß Anlage 1 der Vereinbarung über eine zentrale Arztnummernvergabe nach § 293 Absatz 7 des SGB V jeweils des behandelnden Arztes und des überweisenden Arztes, Art der Behandlung, Beginn und Ende der Behandlung je Quartal, Art der Inanspruchnahme, Entbindungsdatum, Fallkosten inklusive der Dialysesachkosten, Erkrankungs- und Leistungsbereich nach § 116b Absatz 1 SGB V, Diagnoseart, Diagnosen nach der Internationalen Klassifikation der Krankheiten in der jeweiligen vom BfArM im Auftrag des Bundesministeriums für Gesundheit herausgegebenen deutschen Fassung nach § 295 Absatz 1 Satz 2

SGB V mit Angaben zur Diagnosesicherheit, zur Lokalisation und zum Diagnosedatum soweit standardmäßig über §§ 295 und 295a übermittelt, Befunde der Zahnärzte, bei ambulanten Operationen die Operationen- und Prozedurenschlüssel nach§ 295 Absatz 1 Satz 4 SGB V mit Lokalisation der Operation oder Prozedur und Datum, abgerechnete ärztliche und zahnärztliche Gebührenpositionen mit Behandlungsdatum, mit Anzahl der Gebührenpositionen und den in Rechnung gestellten Entgelten, der Abrechnungsbegründung und der Sachkostenbezeichnung, Angabe zu Zweitmeinungen gemäß der Richtlinie über die Konkretisierung des Anspruchs auf eine unabhängige ärztliche Zweitmeinung gemäß § 27b Absatz 2 SGB V, Angaben zur Vermittlungsart durch und Kontaktart mit der Terminservicestelle nach § 87a Absatz 3 Satz 5 Nummer 3 bis 6 SGB V, der Datumsangabe des Kontakts und der Arztgruppe des behandelnden Arztes, Bewertung und die Vertragsnummer für Verträge im Rahmen der hausarztzentrierten Versorgung nach § 73b SGB V und der integrierten Versorgung nach § 140a SGB V,

- nach § 300 SGB V zur Abgabe von Arzneimitteln: Pharmazentralnummer des abgegebenen Arzneimittels einschließlich der nach § 300 Absatz 3 SGB V vereinbarten Sonderkennzeichen, Verordnungsdatum, Betriebsstättennummer, lebenslange Arzt- oder Zahnarztnummer des verordnenden Arztes gemäß Anlage 1 der Vereinbarung nach § 293 Absatz 7 SGB V über eine zentrale Arztnummernvergabe, Datum der Abgabe durch die Apotheke an den Patienten, Kennzeichen der Begründungspflicht für zahnärztliche Verordnung, Vertragskennzeichen für einzelvertragliche Vereinbarungen zwischen Krankenkassen und Leistungserbringern, Institutionskennzeichen der abgebenden Apotheke, Kennzeichnung zum Sitz im In- oder Ausland der Apotheke, Mengenfaktor laut Verordnung je Position, den verwendeten und an den Versicherten abgegebenen Anteil der Packung je Fertigarzneimittel und die Kennzeichnung von Verwurf bei parenteralen Zubereitungen oder bei der Abgabe wirtschaftlicher Einzelmengen, Noctu-Kennzeichen zur Befreiung von der Notfallgebühr, Angaben zu Aut-Idem zur Austauschbarkeit von Wirkstoffen, Wirkstoffverordnung und Dosierung, den verordnungszeilenbezogenen Sozialversicherungs-Bruttobetrag, die gesetzlichen Abschläge von Apotheken und Großhändlern in Eurobeträgen, rezeptbezogene Zuzahlungen des Versicherten in Eurobeträgen, Angaben zur Eigenbeteiligung des Versicherten in Eurobeträgen,

- nach § 301 Absatz 1 SGB V zur stationären Versorgung inklusive der entsprechenden Daten der vor- und nachstationären sowie ambulanten Krankenhausbehandlung, soweit diese Daten in den Richtlinien und Vereinbarungen

für den Datenaustausch der einzelnen Leistungserbringergruppen im Rahmen der Abrechnung mit der Gesetzlichen Krankenversicherung vorgesehen sind: Institutionskennzeichen der behandelnden Einrichtung, Aufnahme- und den Entlassungstag mit Aufnahme- und Entlassungsgrund, Tag der Behandlung, Angaben zur Doppeluntersuchung aus Information aus der Abrechnungsbegründung, Art der Belegleistung, Haupt- und primäre sowie sekundäre Neben-, Aufnahme- und Entlassungsdiagnosen nach der Internationalen Klassifikation der Krankheiten in der jeweiligen vom BfArM im Auftrag des Bundesministeriums für Gesundheit herausgegebenen deutschen Fassung nach § 301 Absatz 2 Satz 1 SGB V mit Lokalisation der Diagnose und Diagnoseart, beteiligte Fachabteilungen mit jeweiligem Aufnahme- und Entlassungsdatum, Leistungsart mit Leistungsschlüssel und Leistungstag, durchgeführte Prozeduren nach dem Operationen- und Prozedurenschlüssel mit Datum und Lokalisation der Operation oder Prozedur, abgerechnete Fallpauschalen nach den diagnosebezogenen Fallgruppen, Angaben zu Entgelten, inklusive der nach § 7 Absatz 1 des Krankenhausentgeltgesetzes für Pflegeleistungen abgerechneten Entgelte, der Entgelte nach der Bundespflegesatzverordnung und der Entgelte zu ambulanten Leistungen sowie der Einzelvergütung mit Erläuterung, Beatmungsstunden, bei ärztlicher Verordnung von Krankenhausbehandlung die Arztnummer des einweisenden Arztes, bei Verlegung das Institutionskennzeichen des veranlassenden Krankenhauses, bei Notfallaufnahme die die Aufnahme veranlassende Stelle,

- nach den §§ 301a und 302 SGB V zur Versorgung mit Heil- und Hilfsmitteln, Versorgung mit Krankentransportleistungen, Versorgung mit häuslicher Krankenpflege, Versorgung mit Hebammenhilfe sowie Versorgung mit digitalen Gesundheitsanwendungen: Abrechnungscode, Tarifkennzeichen, Abrechnungspositionsnummer der erbrachten Einzelleistung oder der abgegebenen Leistung mit Anzahl oder Menge, Einzelbetrag der Abrechnungsposition, Betrag gesetzlicher Zuzahlung, Eigenanteil und Mehrkosten, Betriebsstättennummer des verordnenden Arztes, lebenslange Arztnummer des verordnenden Arztes gemäß Anlage 1 der Vereinbarung nach § 293 Absatz 7 SGB V über eine zentrale Arztnummernvergabe, Institutionskennzeichen des Leistungserbringers, Institutionskennzeichen des Krankenhauses, in dem die Hebamme als Beleghebamme die Leistungen erbracht hat, Datum der Leistungserbringung und der Verordnung, Behandlungsbeginn und Behandlungsende, gefahrene Kilometer der Hebamme, des Entbindungspflegers und des Krankentransports, Diagnose- oder Indikationsschlüssel mit Spezifikation des An-

wendungsortes, Kennzeichnung des Hausbesuchs, Kennzeichen für Hilfsmittel, Kennzeichen für die Verordnungsart bei Heilmitteln und für die Verordnungsbesonderheiten, Inventarnummer für Hilfsmittel im Wiedereinsatz, Pharmazentralnummer, Positionsnummer für Produktbesonderheiten, Geburtsdatum des Kindes oder der errechnete Geburtstermin bei vorgeburtlichen Leistungen, Anzahl der geborenen Kinder,

- Angaben zum Vitalstatus und zum Sterbedatum des Versicherten.

Aus dieser Vielzahl an Daten ergibt sich: Über den bisherigen, auf den Morbi-RSA-bezogenen Datenkranz hinaus werden insbesondere personenbezogene Daten der Versicherten zu den vorgenommenen Behandlungsmethoden und Diagnosen sowie personenbezogene Daten der involvierten Leistungserbringer verarbeitet. Umfasst sind neuerdings auch die Daten zu ambulanten Krankenhausbehandlungen.

Der erweiterte Datenkranz ermöglicht den forschenden Nutzungsberechtigten einen größeren Überblick über die Behandlung und Verläufe von Erkrankungen. Zusammenhänge von Krankheiten und ihre Behandlungen können umfassend untersucht werden.[3] Beispiele potentieller Fragestellungen, die mithilfe des erweiterten Datenkranzes erforscht werden können, sind etwa die Vermeidung von Fehlmedikationen und gefährlichen Wechselwirkungen von Arzneimitteln sowie die Suche nach Ansatzpunkten für geringere Hospitalisierungsraten und eine höhere Lebenserwartung. Aus den Forschungsergebnissen können Rückschlüsse für die Gesundheitsversorgung der Bevölkerung gezogen werden.

b) Datenverarbeitungen bei den gesetzlichen Krankenkassen und dem Spitzenverband Bund der Krankenkassen

Die gesetzlichen Krankenkassen oder von diesen beauftragte Stellen generieren für die Sozialdaten, die sich jeweils auf einen Versicherten beziehen, gemäß § 303b Abs. 1 Satz 1 SGB V ein sog. Lieferpseudonym. Technisch erstellen die Krankenkassen dieses Lieferpseudonym mithilfe eines ihnen vom RKI zuvor übermittelten geheimen Jahresschlüssels und der Versichertennummer des Versicherten. Dabei wenden sie eine sichere Hash-Funktion (z.B. SHA-3-256, gemäß BSI-Empfehlungen) an. Das Lieferpseudonym erlaubt eine eindeutige Identifizierung des jeweiligen Versicherten im Berichtszeitraum. Der Berichtszeitraum erstreckt sich immer auf ein Kalenderjahr; die Daten werden zwar mindestens quartalsweise erhoben, an das FDZ aber nur jährlich geliefert.

---

[3] Referentenentwurf des Bundesministeriums der Gesundheit zur „Verordnung zur Neufassung der Datentransparenzverordnung und zur Änderung der Datentransparenz-Gebührenverordnung“ vom 16.06.2020, S. 30.

Die gesetzlichen Krankenkassen übermitteln die Sozialdaten jedes Versicherten und das entsprechende Lieferpseudonym insoweit an den Spitzenverband Bund der Krankenkassen. Auch mit der Übermittlung der Daten können sie Stellen beauftragen. Nach Bestätigung der erfolgreichen Übermittlung löscht die Krankenkasse den geheimen Jahresschlüssel, so dass sie danach selbst nicht mehr von Versichertennummer auf Lieferpseudonym schließen kann. Die Versichertennummer wird nicht an den Spitzenverband Bund der Krankenkassen mitübermittelt, um eine nachträgliche Identifizierung des Versicherten zusätzlich zu erschweren. Der Spitzenverband Bund der Krankenkassen fungiert als eine Datensammelstelle und erhebt und speichert die Sozialdaten inklusive Lieferpseudonym.

Der Spitzenverband Bund der Krankenkassen führt einerseits die Daten in versichertenbezogene Datensätze zusammen, andererseits pseudonymisiert er die in den Datensätzen enthaltenen Angaben zu den Leistungserbringern. Für die Pseudonymisierung der Daten der Leistungserbringer wählt der Spitzenverband Bund der Krankenkassen gemäß § 303b Absatz 3 Satz 2 SGB V ein Verfahren nach dem Stand der Technik, bei dem die den einzelnen Leistungserbringer identifizierenden Ziffern der Betriebsstättennummer, der lebenslangen Arztnummer und des Institutionskennzeichens durch jeweils ein dauerhaftes Pseudonym ersetzt wird. Die aus den übrigen Ziffern ableitbaren Informationen zu den jeweiligen Leistungserbringern sind im Datensatz gesondert aufzuführen. Dies betrifft bei der Betriebsstättennummer die regionale Zuordnung in Form des Landes- oder Bezirksstellenschlüssels der Kassenärztlichen Vereinigung, bei der lebenslangen Arztnummer die Facharztbezeichnung und beim Institutskennzeichen die Art der Einrichtung oder die Personengruppe sowie die regionale Zuordnung durch Kennzeichnung des Bundeslandes.

Der Spitzenverband Bund der Krankenkassen generiert weiterhin für jeden Einzeldatensatz eine Arbeitsnummer, von der nach dem derzeitigen Stand der Technik keine Rückschlüsse auf das von den Krankenkassen generierte Lieferpseudonym (und das periodenübergreifende, von der Vertrauensstelle generierte Pseudonym nach § 303c Abs. 1 SGB V) gezogen werden können. Damit kann von der Arbeitsnummer auch grundsätzlich nicht auf den jeweiligen Versicherten zurückgeschlossen werden. Jeder Einzeldatensatz wird mit dieser Arbeitsnummer gekennzeichnet.

Der Spitzenverband Bund der Krankenkassen übermittelt die Einzeldatensätze der Versicherten ohne das Lieferpseudonym einschließlich der von ihm pseudonymisierten Daten

der Leistungserbringer und der Arbeitsnummer gemäß § 303b Abs. 3 Nr. 1 SGB V an das FDZ.[4]

Das technische Verfahren der Datenübermittlung ist noch nicht endgültig definiert. Jedenfalls die in der Übergangsregelung des § 12 DaTraV in Bezug genommenen Datenübermittlungen erfolgen über die sichere und etablierte Datenaustauschplattform GKVNet DIC (https://datenaustausch.gkv-spitzenverband.de/#/pages/Home). Für die Registrierung bei dieser Plattform hat das BfArM bereits ein Institutionskennzeichen erhalten, womit ein Zertifikat beim sog. ITSG Trustcenter beantragt wurde. Die DaTraV-Daten werden vor der Übermittlung über die Plattform verschlüsselt und auf einem SFTP-Server bereitgestellt. Der private Schlüssel zur Entschlüsselung der Dateien liegt – zusätzlich durch eine Passphrase abgesichert – in einer gesicherten Umgebung innerhalb der Infrastruktur des FDZ. Ferner muss ein Mitarbeiter des FDZ für den Abruf der verschlüsselten Daten registriert und für das Verfahren auf der Plattform freigeschaltet sein.

c) Datenverarbeitungsvorgänge im FDZ

Nachfolgend werden die im FDZ stattfindenden Datenverarbeitungsvorgänge in den Blick genommen. Diese vollziehen sich im Wesentlichen in drei Phasen: Die Aufbereitung und Analyse der Einzeldatensätze (A.III.3.c)aa)), das Antragsverfahren der Nutzungsberechtigten und das Öffentliche Antragsregister (A.III.3.c)bb)) und die Datenbereitstellung an die Nutzungsberechtigten (A.III.3.c)cc)).

aa) Datenverarbeitungen im Zusammenhang mit der Aufbereitung und Analyse der Einzeldatensätze

Nach Entgegennahme und Speicherung der Einzeldatensätze [siehe A.III.3.c)aa)(1)] und eines periodenübergreifenden Pseudonyms [siehe A.III.3.c)aa)(2)] führt das FDZ die Daten zusammen und bereitet sie intern zum Zwecke der Bereitstellung an die potentiell Nutzungsberechtigten auf [siehe A.III.3.c)aa)(3)].

(1) Entgegennahme und Speicherung der Daten im FDZ

Das FDZ nimmt die DaTraV-Daten, bestehend aus den Sozialdaten der Versicherten und den pseudonymisierten Daten der Leistungserbringer sowie die entsprechenden Arbeitsnummern von dem Spitzenverband Bund der Krankenkassen auf einem Server entgegen

[4] Ausweislich der Begründung zur DaTraV vereinbart das Nähere zur technischen Ausgestaltung dieser Datenübermittlung der Spitzenverband Bund der Krankenkassen mit dem Forschungsdatenzentrum und der Vertrauensstelle gemäß § 303b Absatz 3 Satz 3 SGB V spätestens bis zum 31. Dezember 2021.

und führt dort eine Prüfung auf enthaltene Schadsoftware aus (sog. Datenschleuse). Sodann importiert das FDZ die Daten in seine internen Datenbanksysteme und speichert sie im sog. isolierten DaTraV-IT-System. Dabei handelt es sich um einen Bereich des DaTraV-IT-Systems, der neben IT-Administratoren nur den FDZ-Analysten, jedoch nicht den Nutzungsberechtigten zugänglich ist. Unter anderem wird dadurch dem insgesamt hohen Schutzbedarf der DaTraV-Daten Rechnung getragen.

(2) Pseudonymisierung bei der Vertrauensstelle und Empfang der Pseudonyme

Parallel zu der Übermittlung der Einzeldatensätze an das FDZ übermittelt der Spitzenverband Bund der Krankenkassen Daten an die beim RKI errichtete Vertrauensstelle. Bei diesen Daten handelt es sich gemäß § 303b Abs. 3 Nr. 2 SGB V um eine Liste mit den von den Krankenkassen generierten Lieferpseudonymen einschließlich der Arbeitsnummern, die zu den an das FDZ übermittelten Einzeldatensätzen für das jeweilige Lieferpseudonym gehören.

Die Vertrauensstelle überführt diese Lieferpseudonyme nach einem einheitlich anzuwendenden Verfahren (dazu sogleich) in perioden- und krankenkassenübergreifende Pseudonyme. Diese sog. permanenten Pseudonyme gewährleisten also, dass der einem konkreten Versicherten zugewiesene Einzeldatensatz nicht nur über längere Zeiträume, sondern auch trotz eines zwischenzeitlichen Wechsels der Krankenkasse nachverfolgbar bleibt.[5] Diese langfristige Nachverfolgbarkeit ist erforderlich, um etwa in der epidemiologischen Forschung Längsschnittanalysen durchführen zu können.

Das Verfahren zur Erzeugung des Pseudonyms muss dem jeweiligen Stand der Technik und Wissenschaft entsprechen. Es ist so zu gestalten, dass für das jeweilige Lieferpseudonym des Einzeldatensatzes eines jeden Versicherten bundesweit eindeutig und unabhängig von seiner Krankenkassenzugehörigkeit jeweils dasselbe periodenübergreifende Pseudonym erstellt wird.

Die Pseudonymisierungssoftware wird vom BSI zur Verfügung gestellt und kann durch das RKI in ihrer Funktionalität nicht geändert werden. [redacted] Ist dieses Lieferpseudonym tatsächlich in den vom Spitzenverband Bund der Krankenkassen gelieferten Daten enthalten,

5 Vgl. Kühling/Schildbach, Rechtsgutachten über den sozialdatenschutzrechtlichen Weiterentwicklungsbedarf im SGB V und SGB X im Hinblick auf Big-Data-Anwendungen vom 04.09.2019, S. 53.

identifiziert die Pseudonymisierungssoftware für einen kurzen Moment im Hauptspeicher des Rechners die tatsächliche Versicherungsnummer. Letztere gehört zu einer konkreten Arbeitsnummer. Gleichzeitig verwendet die Software einen geheimen Master-Schlüssel, der nur dem RKI als Vertrauensstelle bekannt ist, um das permanente Pseudonym zu erzeugen. Das RKI hat – abgesehen von dem kurzen Moment während der Erzeugung des Pseudonyms – keine Kenntnis von den Versichertendaten.

Daraus folgt, dass aus dem periodenübergreifenden Pseudonym ohne Weiteres weder auf das Lieferpseudonym (1), noch auf die Versichertennummer (2) und somit auch nicht auf die Identität des Versicherten geschlossen werden kann.

Nach erfolgter Pseudonymisierung übermittelt die Vertrauensstelle dem FDZ verschlüsselt und über ein sicheres Übermittlungsverfahren die Liste der periodenübergreifenden, permanenten Pseudonyme mit den dazugehörigen Arbeitsnummern. Unmittelbar nach der erfolgreichen Übermittlung löscht die Vertrauensstelle die den periodenübergreifenden Pseudonymen zugrundeliegenden Lieferpseudonyme und Arbeitsnummern sowie die periodenübergreifenden Pseudonyme.

Das FDZ empfängt die Liste mit den periodenübergreifenden, permanenten Pseudonymen und den dazugehörigen Arbeitsnummern, prüft sie in der Datenschleuse auf Schadsoftware und überführt sie erst danach in das interne Datenbanksystem zur Speicherung im isolierten DaTraV-IT-System. Anschließend gleicht das FDZ die Arbeitsnummern auf dieser Liste mit den Arbeitsnummern ab, die bereits im isolierten DaTraV-IT-System in den Einzeldatensätzen der Versicherten gespeichert sind. Anhand dieses Abgleichs führt es das periodenübergreifende, permanente Pseudonym mit dem jeweiligen Einzeldatensatz des Versicherten zusammen und speichert es zusammen mit diesem Einzeldatensatz im isolierten DaTraV-IT-System. Das FDZ kann nun auf Basis dieser Zusammenführung den vom Spitzenverband Bund der Krankenkassen mit der Arbeitsnummer gelieferten Datensatz dem bisherigen Datenbestand des pseudonymen Versicherten zuordnen, so dass sich dessen periodenübergreifende Krankengeschichte ergibt. Die so generierten Einzeldatensätze eines jeden Versicherten werden durch das FDZ jährlich um die aktualisierten Daten dieses pseudonymen Versicherten ergänzt.

Das FDZ hat die versichertenbezogenen Einzeldatensätze spätestens nach 30 Jahren zu löschen.

(3) Interne Aufbereitung der Daten

Dieser im isolierten DaTraV-IT-System gespeicherte DaTraV-Datenbestand wird primär dazu genutzt, um die Auswertungen der nutzungsberechtigten Forscher durchzuführen.

Die zu diesem Zweck erforderlichen Auswertungsprogramme werden ausschließlich von den Forschenden selbst erstellt und entwickelt. Die gespeicherten Originaldaten werden im isolierten DaTraV-IT-System vom FDZ ausschließlich auf Grundlage dieser Programme ausgewertet.

Das FDZ beschränkt sich insoweit auf die Beratung und Schulung und erstellt verschiedene aus den DaTraV-Daten aufbereitete Standardauswertungen. Diese werden pauschal und nicht mit Blick auf ein bestimmtes Forschungsvorhaben geschaffen. Solche Standardauswertungen werden von den FDZ-Analysten – teilweise in Abstimmung mit den nutzungsberechtigten Forschern – im isolierten DaTraV-IT-System erstellt. Sie umfassen Scientific Use Files, Referenzauswertungen und Datencharakterisierungen.

- Bei Scientific Use Files handelt es sich um Daten, die so erzeugt werden, dass bestimmte inhaltliche Merkmale des originalen Datenbestandes in größerem Umfang enthalten sind.

- Referenzauswertungen sind standardisierte, abstrakt gehaltene Auswertungen, die im isolierten DaTraV-IT-System aus vergangenen Forschungsprojekten vorliegen.

- Datencharakterisierungen sind inhaltliche und technische Datensatzbeschreibungen, die die Daten strukturell (hinsichtlich Datenfelder und -typen) als auch inhaltlich (beispielsweise hinsichtlich Häufigkeiten, Anzahl der Versicherten etc.) beschreiben und lediglich einer groben Plausibilitätskontrolle dienen.

Anhand dieser unterschiedlichen Auswertungen kann das FDZ standardisierte Datensätze generieren.

Ferner stellt das FDZ sog. Hilfstabellen bereit, insbesondere zur Zuordnung von Wirkstoffcodes (ATC) zu den Arzneimittelkennzeichen nach § 300 Absatz 3 SGB V (PZN), oder die Auswahl von Diagnosen über Strukturinformationen aus dem ICD-Katalog.

Weiterhin werden sog. Basisauswertungspopulationen erstellt, die sich ergeben, wenn definierte Auswahlkriterien angewendet werden (z.B. folgende Kriterien: „Nur Versicherte mit konsistenten Angaben zu Geschlecht und Geburtsjahr über alle beantragten Berichtsjahre hinweg“ oder „Nur Versicherte ohne Versichertentage im Ausland“).

Darüber hinaus erstellt das Forschungsdatenzentrum vollumfänglich anonymisierte Datensätze (sog. Public Use Files) und macht diese öffentlich in seinem Web-Portal verfügbar, insbesondere zu Schulungs-, Entwicklungs- und Testzwecken. Es fördert mit jenen Maßnahmen schließlich die wissenschaftliche Erschließung der Daten.

Das FDZ prüft und sichert sukzessive die Qualität der Einzeldatensätze und unterstützt die Weiterentwicklung der Methodik.

bb) Datenverarbeitungen im Zusammenhang mit dem Antragsverfahren und dem Öffentlichen Antragsregister

(1) Antragstellung

Nutzungsberechtigt hinsichtlich der durch das FDZ aufbereiteten Daten sind nach § 303e Abs. 1 SGB V ausschließlich die folgenden Institutionen: Spitzenverband Bund der Krankenkassen, Bundes- und Landesverbände der Krankenkassen, Krankenkassen, Kassenärztliche Bundesvereinigungen und Kassenärztliche Vereinigungen, für die Wahrnehmung der wirtschaftlichen Interessen gebildete maßgebliche Spitzenorganisationen der Leistungserbringer auf Bundesebene, Institutionen der Gesundheitsberichterstattung des Bundes und der Länder, Institutionen der Gesundheitsversorgungsforschung, Hochschulen, nach landesrechtlichen Vorschriften anerkannte Hochschulkliniken, öffentlich geförderte außeruniversitäre Forschungseinrichtungen und sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung, sofern die Daten wissenschaftlichen Vorhaben dienen, Gemeinsamer Bundesausschuss, das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, das Institut des Bewertungsausschusses, Beauftragter der Bundesregierung und der Landesregierungen für die Belange der Patientinnen und Patienten, die für die Wahrnehmung der Interessen der Patientinnen und Patienten und der Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen auf Bundesebene, das Institut für Qualitätssicherung und Transparenz im Gesundheitswesen, das Institut für das Entgeltsystem im Krankenhaus, die für die gesetzliche Krankenversicherung zuständigen obersten Bundes- und Landesbehörden und deren jeweilige nachgeordnete Bereiche sowie die übrigen obersten Bundesbehörden, die Bundesärztekammer, die Bundeszahnärztekammer, die Bundespsychotherapeutenkammer sowie die Bundesapothekerkammer und die Deutschen Krankenhausgesellschaft. Das BfArM gehört nicht zu den Nutzungsberechtigten, solange es die Aufgaben des FDZ wahrnimmt. Eine Antragstellung des BfArM ist somit ausgeschlossen.

Für das FDZ gilt innerhalb des BfArM die Pflicht, Maßnahmen zur Gewährleistung der organisatorischen Selbstständigkeit zu treffen. Dies betrifft die räumliche Trennung von den übrigen Einheiten des BfArM, die eigenständige Verwahrung der DaTraV-Daten ausschließlich in der vorgesehenen separaten IT-Infrastruktur, auf die innerhalb des BfArM nur Beschäftigte des FDZ Zugriff haben, und die organisatorische Trennung durch die Einrichtung einer getrennten Organisationseinheit.

Die o.g. Nutzungsberechtigten beantragen gemäß § 7 Abs. 1 DaTraV Zugang zu den im FDZ vorhandenen Daten über eine elektronisch bereitgestellte Antragsverwaltung. In dieser Antragsverwaltung hat der Nutzungsberechtigte die folgenden Informationen anzugeben und mittels elektronischem Antragsformular an das FDZ zu übermitteln:

- seinen Namen und die Anschrift,
- den Zweck der Datenverarbeitung (Nutzungszweck),
- den methodischen Ansatz der Datenverarbeitung,
- den Umfang und die Struktur der beantragten Daten sowie eine nachvollziehbare Darlegung, dass Umfang und Struktur der beantragten Daten geeignet und erforderlich sind, um die zu untersuchende Frage zu beantworten,
- die an der Datenverarbeitung beteiligten Personen und
- ob eine Zusammenführung der beantragten Daten mit anderen Datenbeständen vorgesehen ist.

Wenn der Nutzungsberechtigte eine krankenkassenbezogene Aufbereitung der Daten beantragt, hat er dem Antrag zusätzlich die Einwilligung der betroffenen Krankenkassen beizufügen.

Der Antragsteller verpflichtet sich bei der Datenverarbeitung darauf zu achten, keinen Bezug zu Personen, Leistungserbringern oder Leistungsträgern herzustellen, geeignete technische und organisatorische Maßnahmen umzusetzen und die Rechte der betroffenen Personen zu wahren. Zudem sind die Daten nur für die Zwecke zu nutzen, für die sie zugänglich gemacht wurden.

(2) Öffentliches Antragsregister

Das FDZ baut gemäß § 303d Abs. 1 Nr. 6 SGB V i.V.m. § 9 DaTraV ein öffentliches Antragsregister mit Informationen zu den antragstellenden Nutzungsberechtigten und zu den Vorhaben auf, für die die Daten beantragt wurden. Das öffentliche Antragsregister ist für jedermann online abrufbar. Darin ist für jeden Antrag Folgendes anzugeben:

- Name und Anschrift des Nutzungsberechtigten,
- der oder die Zwecke nach § 303e Absatz 2 SGB V,
- der Titel des Vorhabens sowie eine kurze Beschreibung des Vorhabens und des mit dem Vorhaben verfolgten Forschungsziels,

- eine kurze Ergebnisdarstellung nach Veröffentlichung von Ergebnissen oder Verweise auf die Publikationen, die auf den Ergebnissen des Vorhabens beruhen und das Jahr der Entscheidung über den Antrag.

Mit Einwilligung der betroffenen Nutzungsberechtigten können darüber hinausgehende Angaben zum Antrag in das Antragsregister aufgenommen werden. Auch Angaben zu beteiligten Personen bedürfen einer Einwilligung, bevor eine Aufnahme im öffentlichen Antragsregister erfolgt.

(3) Antragsprüfung durch das FDZ

Die Antragsbearbeitung erfolgt (teil-) automatisiert. Das FDZ – namentlich die FDZ-Analysten – prüfen die Anträge zur Datenverarbeitung innerhalb des Antragsverwaltungssystems gemäß § 7 Abs. 1 DaTraV dahingehend, ob

- der Antragsteller nach § 303e Absatz 1 SGB V zur Verarbeitung berechtigt ist (Nutzungsberechtigter),
- der angegebene Nutzungszweck mindestens einem der in § 303e Absatz 2 SGB V aufgeführten Nutzungszwecken entspricht,
- die Datenverarbeitung nach § 303e Absatz 2 SGB V jeweils für die Erfüllung von Aufgaben des Nutzungsberechtigten erforderlich ist,
- im Antrag nachvollziehbar dargelegt ist, dass der Umfang und die Struktur der beantragten Daten geeignet und erforderlich sind, um die zu untersuchende Frage zu beantworten,
- das im Antrag angegebene Vorhaben mit den beim FDZ vorliegenden Daten bearbeitet werden kann und
- die Verpflichtungserklärung des Antragstellers nach § 7 Abs. 3 DaTraV vorliegt.

In praktischer Hinsicht bedeutet dies eine Prüfung, ob die vom antragstellenden Nutzungsberechtigten gewählte Auswertungsmethodik nur auf die Daten zugreift, die für die Beantwortung der konkreten Forschungsfragen benötigt werden. Zu diesem Zweck bewerten die FDZ-Analysten, ob die gewählte Auswertungsmethodik generell geeignet scheint, das Forschungsziel zu erreichen.

Vor dem Hintergrund entscheidet das FDZ nach pflichtgemäßem Ermessen über die Art der Bereitstellung der Daten nach § 303e Abs. 3 und 4 SGB V und § 10 Abs. 1 DaTraV (vgl. zu den Arten der Datenbereitstellung sogleich unter A.III.3.c)cc)).

Das FDZ entscheidet über den Antrag durch Verwaltungsakt und übersendet dem Antragsteller den Bescheid schriftlich und/oder elektronisch. Der Verwaltungsakt kann mit Auflagen verbunden werden, insbesondere mit der Auflage, die vorgesehene Zusammenführung der beantragten Daten mit externen Datenbeständen zu unterlassen.

Die im Rahmen der Antragstellung von den Nutzungsberechtigten erhobenen personenbezogenen Daten werden im Rahmen eines Institutions- und Nutzerregisters in der IT-Systeminfrastruktur gespeichert. Dauerhaft werden dort nur die Daten mit Bezug zu Institutionen gespeichert, deren Antrag positiv beschieden wurde. Ein Löschkonzept für die Daten der übrigen Institutionen wird noch zu erstellen sein.

cc) Bereitstellung der Daten des FDZ an die Nutzungsberechtigten

(1) Stufen der Bereitstellung der Daten und Prüfung der Ergebnismengen

Die Daten können den Nutzungsberechtigten grundsätzlich in drei (in § 10 Abs. 1 Nr. 1-3 DaTraV näher beschriebenen) Stufen zugänglich gemacht werden. Welche Stufen bis zur Bereitstellung einer Ergebnismenge im Einzelfall zu durchlaufen sind, hängt insbesondere von dem Forschungsziel des Nutzungsberechtigten ab.

- Die Bereitstellung kann in Gestalt von Ergebnismengen erfolgen, die auf Standardauswertungen beruhen und in aggregierter und anonymisierter Form im FDZ vorliegen. Der Nutzungsberechtigte erhält nur die aggregierten und anonymisierten Ergebnismengen, ohne ein Auswertungsprogramm entwickeln und einreichen zu müssen. Teilweise sind Standardauswertungen, in Gestalt der Public Use Files, sogar im Webportal des FDZ ohne Registrierung verfügbar.

- Weiterhin können dem Nutzungsberechtigten aggregierte und anonymisierte Ergebnismengen zur Verfügung gestellt werden, die auf vom Nutzungsberechtigten selbst erstellten Auswertungsprogrammen beruhen. Kommt das FDZ im Zuge der Antragsprüfung zu dem Ergebnis, dass es für die Bereitstellung der Daten eines Auswertungsprogrammes bedarf, teilt es dies dem Nutzungsberechtigten vor Bewilligung des Antrags mit und fordert den Antragsteller auf, ein Auswertungsprogramm vorzulegen. Zur Ausarbeitung und Prüfung geeigneter Auswertungsprogramme werden dem Nutzungsberechtigten regelmäßig Testdatensätze und für die Ausarbeitung erforderliche vorläufige Auswertungen und Zwischenergebnisse bereitgestellt. Dabei handelt es sich um projektspezifische synthetische Daten, deren Detailausprägung weitgehend vom konkreten Forschungsvorhaben abhängt. Mithilfe der synthetischen Daten muss die Fragestellung hinreichend genau beantwortet werden können, um die Auswertungen mit diesen Daten bis zu einem fortgeschrittenen Entwicklungsstadium führen zu können.

Die Nutzungsberechtigten erstellen auf dieser Grundlage das Auswertungsprogramm und reichen es beim FDZ ein. Mit dem Auswertungsprogramm des Nutzungsberechtigten wertet das FDZ die Originaldaten aus und übermittelt die Ergebnisse in Form von aggregierten Ergebnismengen an die Nutzungsberechtigten.

– Das FDZ kann einem Nutzungsberechtigten auch pseudonymisierte Einzeldatensätze zugänglich machen, wenn der antragstellende Nutzungsberechtigte nachvollziehbar darlegt, dass die Nutzung der pseudonymisierten Einzeldatensätze für einen zulässigen Nutzungszweck, insbesondere für die Durchführung seines Forschungsvorhabens und regelmäßig zur Erstellung des Auswertungsprogramms, erforderlich ist. Das FDZ stellt einem Nutzungsberechtigten die pseudonymisierten Einzeldatensätze ohne Sichtbarmachung der Pseudonyme für die Verarbeitung unter Kontrolle des FDZ bereit.

  Das FDZ stellt sicher, dass diese Daten nur solchen Personen bereitgestellt werden, die einer Geheimhaltungspflicht nach § 203 des Strafgesetzbuches unterliegen oder wenn sie vor dem Zugang gemäß § 1 Abs. 2, 3 und 4 Nr. 2 des Verpflichtungsgesetzes zur Geheimhaltung verpflichtet wurden. Ferner legt das FDZ im Einvernehmen mit dem BSI die erforderlichen spezifischen, technischen und organisatorischen Maßnahmen fest, um die Datenverarbeitung durch den Nutzungsberechtigten auf das erforderliche Maß zu beschränken und um das Risiko einer Identifizierung einzelner Betroffener zu minimieren. Insbesondere ein Kopieren oder sonstiges Vervielfältigen der Einzeldatensätze durch die Nutzungsberechtigten wird mithilfe dieser Maßnahmen verhindert. Einzeldatensätze werden nicht an Nutzungsberechtigte übermittelt. Die pseudonymisierten Einzeldatensätze werden lediglich in gesicherter physischer oder virtueller Umgebung bereitgestellt. Geplant ist derzeit die Bereitstellung eines gesicherten Fernzugriffs über eine Analyseplattform (dazu sogleich unter A.III.3.c)cc)(2)). Die konkret anzuwenden Verfahren und Anforderungen sind noch zu definieren. Bei deren Entwicklung, Erprobung und Festlegung ist das BSI einzubeziehen, um ausreichende technische Sicherheit zu gewährleisten. Der Zugang zu pseudonymisierten Einzeldatensätzen durch Nutzungsberechtigte soll auf 30 Arbeitstage pro Antrag begrenzt werden.

Das FDZ betrachtet anschließend die im isolierten DaTraV-IT-System erzeugten Ergebnismengen, bevor es sie an die Nutzungsberechtigten übermittelt. Es prüft, ob mittels der Ergebnismengen oder durch eine beabsichtigte Zusammenführung („Verschneidung“) der Ergebnismengen mit anderen Datenbeständen die Versicherten wieder identifiziert werden können. Die genauen Kriterien und Tiefe der Prüfung werden derzeit eruiert und sind noch im Einzelnen festzulegen.

Kommt das FDZ zu dem Ergebnis, dass eine Re-Identifizierung möglich ist, so wird dieses Risiko durch geeignete Maßnahmen unter Wahrung des angestrebten wissenschaftlichen Nutzens minimiert. Insbesondere wird das FDZ seine an die Ergebnismengen gestellten Anforderungen derart verändern (zum Beispiel durch Heraufsetzen der Mindestfallzahlen einzelner Ergebnisfelder oder deren Streichung), dass Re-Identifizierungen möglichst ausgeschlossen sind.

(2) Voraussichtliche Funktionalität der Systemarchitektur zur Übermittlung der Ergebnismengen

Sodann übermittelt das FDZ dem Nutzungsberechtigten die geprüften Ergebnismengen. Die hierfür vorgesehene Systemarchitektur beim FDZ wird die grundlegenden Aspekte von Datenschutz und Datensicherheit beachten. Ihre Umsetzung wird voraussichtlich erst im Rahmen der zweiten Ausbaustufe realisiert werden. Vorgesehen ist derzeit die im Folgenden skizzierte Funktionalität:

Neben dem isolierten DaTraV-IT-System, auf das nur die FDZ-Analysten Zugriff haben, ist als weitere Komponente der Systemarchitektur des FDZ ein sog. integriertes System vorgesehen, das den Bedürfnissen der nutzungsberechtigten Forscher nach effizienten und möglichst eigenständigen Arbeitsweisen gerecht werden soll. Das integrierte System wird zur Nutzung durch die Forscher selbst konzipiert. Es soll ihnen je Forschungsprojekt einen separaten, abgegrenzten und technisch geschützten Analyseraum zur Verfügung stellen, um die Auswertungsprogramme entwickeln zu können. Jeder Analyseraum soll Analysewerkzeuge, Datenspeicher, ein Dateiablagesystem und ein Versionenkontrollsystem enthalten. In den jeweiligen Analyseraum des integrierten Systems sollen aus dem isolierten DaTraV-IT-System sowohl Standardauswertungen als auch – abhängig vom konkreten Forschungsprojekt – synthetische Daten oder andere ausgewählte Datenausschnitte übertragen werden. Ferner ist derzeit vorgesehen, dem nutzungsberechtigten Forscher, soweit für das konkrete Forschungsvorhaben erforderlich, im Ausnahmefall pseudonymisierte Datensätze im Analyseraum des integrierten Systems zugänglich zu machen.

Je nach Schutzbedarf der in das integrierte System übermittelten Daten sollen unterschiedliche im Einzelnen noch zu definierende Schutzmaßnahmen implementiert werden. In Planung ist etwa, verschiedene Zugriffsrechte an nutzungsberechtigte Personen zu vergeben. Dabei soll sichergestellt werden, dass die übermittelten Daten den jeweiligen Analyseraum auf technischem Wege nicht verlassen können. Weiterhin wird gewährleistet, dass die Übernahme von Daten und Dateien in die Analyseräume von außen nur über eine Datenschleuse erfolgt. Die Datenschleuse überprüft die importierten Daten auf Schadsoftware.

Die Übermittlung von Daten aus dem isolierten DaTraV-IT-System an das integrierte System wird durch Forscher auch *vice versa* erfolgen können. Hat etwa der Forscher sein Auswertungsprogramm im Analyseraum fertiggestellt, kann er es in das isolierte DaTraV-IT-System übertragen und dort, wie oben unter A.III.3.c)aa)(3) beschrieben, ausführen lassen. Ziel der Systemfunktionalität des FDZ ist, dass der nutzungsberechtigte Forscher die Ausarbeitung seiner Auswertungsprogramme in dem im integrierten System befindlichen Analyseraum soweit vorantreiben kann, dass im Idealfall nur eine einzige Iteration im isolierten System erforderlich ist und insofern in jedem Projekt möglichst wenige Datenflüsse zwischen beiden Systemen stattfinden müssen. Zwischen den beiden Systemen herrscht stets eine strikte Trennung. Die dennoch notwendigen und zulässigen Datenflüsse sind noch nicht abschließend definiert und werden im Detail mit dem BSI abgestimmt und unter seiner Beteiligung ausgearbeitet.

Der Nutzungsberechtige darf sämtliche DaTraV-Daten grundsätzlich nicht an Dritte weitergeben. Das FDZ darf auf Antrag der Nutzungsberechtigten ausnahmsweise eine Weitergabe an einen Dritten im Rahmen eines zulässigen Nutzungszwecks genehmigen. Letzteres mag einer Beratung durch Dritte bei der Auswertung von Daten dienen.

Grundsätzlich gilt: Die Nutzungsberechtigten haben dem FDZ zu melden, wenn ein Bezug zu Personen, Leistungserbringern oder Leistungsträgern unbeabsichtigt hergestellt wird. Zudem ist gemäß § 397 Abs. 1 SGB V sowohl die unrechtmäßige Weitergabe der bereitgestellten Daten an Dritte (Nr. 1), als auch jede Verarbeitung zum Zwecke der Herstellung eines Personenbezugs, der Identifizierung von Leistungserbringern oder Leistungsträgern und der bewussten Verschaffung von Kenntnissen über fremde Betriebs- und Geschäftsgeheimnisse (Nr. 2) strafbewehrt.

IV. Notwendigkeit und Verhältnismäßigkeit der Verarbeitungsvorgänge in Bezug auf die Zwecke, Art. 35 Abs. 7 lit. b DS-GVO

Mit der Weiterentwicklung des Datentransparenzverfahrens und des neu definierten Datenkranzes wird die Attraktivität des FDZ erhöht. Die damit einhergehenden Datenverarbeitungsvorgänge sind notwendig und verhältnismäßig im Sinne des Art. 35 Abs. 7 lit. b DS-GVO.

Die vor der Gesetzesänderung bisher genutzte technische Infrastruktur wird die zu erwartende erhöhte Nachfrage und die Anforderungen an ein modernes Forschungsdatenzentrum nicht bewältigen können.[6]

---

[6] Referentenentwurf des Bundesministeriums der Gesundheit zur „Verordnung zur Neufassung der Datentransparenzverordnung und zur Änderung der Datentransparenz-Gebührenverordnung“ vom 16.06.2020, S. 24.

Soweit die Datenverarbeitungen im FDZ bereits feststehen, sind sie geeignet, erforderlich und verhältnismäßig im engeren Sinne. Das gilt sowohl mit Blick auf die personenbezogenen Daten der gesetzlich Versicherten als auch der Leistungserbringer und der nutzungsberechtigten Forscher.

1. Übergeordnete Zwecke der Datenverarbeitungen

Die Verarbeitungen personenbezogener Daten im Zusammenhang mit dem Datentransparenzverfahren bezwecken ausweislich des § 303e Abs. 2 DS-GVO im Allgemeinen die Gewährleistung und Verbesserung der Gesundheitsvorsorge sowie der gesundheitlichen Versorgung. Dies sind überragend wichtige Gemeinschaftsgüter. Ferner wird die Förderung der wissenschaftlichen Forschung bezweckt, die nach den Wertungen der DS-GVO einen eigens zu fördernden Zweck darstellt; Datenverarbeitungen, die der wissenschaftlichen Forschung dienen, sind in vielerlei Hinsicht privilegiert.

Der demokratisch legitimierte Gesetzgeber hat sich dafür entschieden, diese Zwecke mithilfe des in den §§ 303a ff. SGB V geregelten Verfahrens zu erreichen bzw. zu fördern. Hierbei hat er den beteiligten Akteuren – insbesondere dem FDZ – strenge Vorgaben gemacht. Dies umfasst, dass etwa der zu verarbeitende Datenkranz sowie die meisten Verarbeitungsvorgänge hinsichtlich des „Ob“ und des „Wie“ gesetzlich vorgegeben sind. Insofern verbleibt – wie die nachfolgenden Ausführungen zeigen – dem FDZ hinsichtlich der konkreten Umsetzung der Verarbeitungen personenbezogener Daten, insbesondere im Hinblick auf die Daten der gesetzlich Versicherten, lediglich ein restriktiver Umsetzungsspielraum.

2. Einzelne Datenverarbeitungen im FDZ gemäß Art. 4 Nr. 2 DS-GVO

Das FDZ erhält einerseits die pseudonymisierten Daten der Versicherten, die pseudonymisierten Daten der Leistungserbringer und die Arbeitsnummern vom GKV-Spitzenverband sowie andererseits die periodenübergreifenden, permanenten Pseudonyme mit den dazugehörigen Arbeitsnummern von der Vertrauensstelle. Hierin liegt eine Erhebung (und kurzfristige Speicherung) der Daten.

Anschließend gleicht das FDZ die Arbeitsnummern zum Zwecke der Zusammenführung von DaTraV-Daten und Pseudonymen miteinander ab. Nach der Zusammenführung speichert das FDZ die Daten im isolierten DaTraV-IT-System längstens für bis zu 30 Jahre.

Den Datenbestand nutzt das FDZ primär, um die Auswertungen der nutzungsberechtigten Forscher entsprechend des jeweiligen Antrags durchzuführen. Hierin liegen unter III.3.c)aa) im Einzelnen näher beschriebene Verwendungen und Anpassungen der Daten im Sinne des Art. 4 Nr. 2 DS-GVO. Hiermit setzt das FDZ seine in der Rechtsgrundlage des § 303e Abs. 3 SGB V

statuierte Pflicht um. Dementsprechend werden die Daten teilweise, insbesondere im Zuge der Erstellung von Standardauswertungen, anonymisiert.

Schließlich stellt das FDZ den Nutzungsberechtigten Daten in Umsetzung der Rechtsgrundlage des § 10 Abs. 1 DaTraV bereit. Hierbei handelt es sich überwiegend nicht mehr um personenbezogene Daten. Soweit die bereitgestellten Daten noch personenbezogen – d.h. konkret pseudonymisiert – sind, liegt hierin eine Offenlegung an die nutzungsberechtigte Organisation. Letzterer untersteht der jeweilige Forschende, der Zugriff auf die Daten erhält, gemäß Art. 29 DS-GVO.

Im Rahmen des Antragsverfahren und der Antragsprüfung verarbeitet das FDZ ggf. Daten mit Bezug zu in den nutzungsberechtigten Forschungseinrichtungen tätigen natürlichen Personen. Deren Daten werden erhoben, zum Zweck des Antragsverfahren und der Antragsprüfung verwendet und möglicherweise in Umsetzung der Pflicht zur Führung eines öffentlichen Antragsregisters gemäß § 303d Abs. 1 Nr. 6 i.V.m. § 9 DaTraV und nach Einholung einer Einwilligung der betroffenen Personen offengelegt.

3. Geeignetheit

Sämtliche vorstehenden Verarbeitungen der DaTraV-Daten beim FDZ sind geeignet, da sie die unter 1. genannten Zwecke fördern. Alle Verarbeitungen personenbezogener Daten dienen unmittelbar den vom Gesetzgeber vorgegebenen Zielen der Gesundheitsvorsorge, der gesundheitliche Versorgung und der wissenschaftlichen Forschung. Die Gewährleistung hoher Qualitäts- und Sicherheitsstandards bei der Gesundheitsversorgung kann jedenfalls aufgrund des Zugangs zu den DaTraV-Daten unterstützt werden. Das FDZ nutzt die Daten nicht zu anderen Zwecken.

Auch die Verarbeitung der Daten mit Bezug zu in den nutzungsberechtigten Forschungseinrichtungen tätigen Personen ist geeignet, ihre Anträge zu bearbeiten und zu prüfen und ihnen schließlich im Falle positiver Bescheidung die DaTraV-Daten zugänglich zu machen. Soweit personenbezogene Daten einzelner Forscher auch im öffentlichen Antragsregister veröffentlicht werden, ist dies geeignet, um Transparenz und Kontrolle im öffentlichen Interesse zu gewährleisten.

4. Erforderlichkeit

Dem Grundsatz der Datenminimierung nach Artikel 5 Absatz 1 lit. c DS-GVO und dem Merkmal der Erforderlichkeit in den Rechtsgrundlagen der §§ 303a bis 303f SGB V wird darüber hinaus Rechnung getragen. Die Datenverarbeitungen sind erforderlich, wenn sämtliche Zwecke nicht genauso gut mit einem „weniger intensiven“ Mittel erreicht werden können. Weniger in-

tensiv ist ein Datenverarbeitungsvorgang, der das Unionsgrundrecht auf Schutz personenbezogener Daten der betroffenen Person gar nicht oder weniger beeinträchtigt. Dies ist sowohl in quantitativer als auch in qualitativer sowie in zeitlicher Hinsicht zu verstehen.

Die unter der datenschutzrechtlichen Verantwortlichkeit des BfArM im FDZ erfolgenden Datenverarbeitungsvorgänge sind unter jedem dieser Aspekte zur Umsetzung der gesetzlich vorgeschriebenen Pflichten gemäß §§ 303a ff. SGB V und mithin zur Forschung über die Gesundheitsversorgung der Versicherten zwingend erforderlich. Insbesondere ist in quantitativer Hinsicht die Verarbeitung der DaTraV-Daten, die über den bisherigen, auf den Morbi-RSA der Krankenkassen fokussierten Datenkranz hinausgehen, erforderlich. Denn die Erweiterung des Datenkranzes ermöglicht den forschenden Nutzungsberechtigten im Vergleich zum bisherigen Datenkranz einen Überblick über die Behandlung und Verläufe von Erkrankungen. Nur auf diese Weise können Zusammenhänge von Krankheiten und ihre Behandlungen untersucht und Rückschlüsse für die Gesundheitsversorgung der Bevölkerung gezogen werden. Ein kleinerer Datenkranz könnte die Repräsentativität der wissenschaftlichen Auswertungen beeinträchtigen und wäre mithin schon nicht gleich geeignet, um die Zwecke zu erreichen. Das BfArM verarbeitet in quantitativer Hinsicht die DaTraV-Daten im Sinne der § 303b Abs. 1 Satz 1 SGB V, § 3 Abs. 1 DaTraV, die ihm vom GKV-Spitzenverband und der Vertrauensstelle übermittelt werden.

Bei den DaTraV-Daten handelt es sich um den Krankenkassen vorliegende Stammdaten der Versicherten und um Daten, die den Krankenkassen aus dem Datenaustausch mit den Leistungserbringern zum Zwecke der Abrechnung ohnehin vorliegen. Die Krankenkassen erheben keine neuen, zusätzlichen Daten. Eine solche Sekundärnutzung zu Forschungszwecken wird durch Art. 5 Abs. 1 lit. b in Verbindung mit Art. 89 DS-GVO ausdrücklich privilegiert. Jenseits der Daten, die gesetzlich vorgegeben sind, erhebt das FDZ keine personenbezogenen Daten.

Die Datenverarbeitungsvorgänge sind auch in qualitativer Hinsicht erforderlich, da keine andere Art der Datenverarbeitung ersichtlich ist, die die Zwecke gleichermaßen effektiv erreicht, und zugleich weniger intensiv in die Grundrechte der betroffenen Personen eingreift. Es wird, soweit sie bereits definiert ist, eine datensparsame Systemarchitektur eingesetzt. Durch die Verwendung des mithilfe der Pseudonymisierungssoftware des BSI erstellten permanenten Pseudonyms werden die Daten der gesetzlich Versicherten so verarbeitet, dass ohne Weiteres kein Rückschluss auf ihre Person möglich ist. Gleiches gilt für die personenbezogenen Daten der Leistungserbringer, die ebenfalls ausschließlich in pseudonymisierter Form verarbeitet werden. Jedes Mal, bevor eine Ergebnismenge nutzungsberechtigten Forschern über die IT-Infrastruktur des FDZ zugänglich gemacht wird, erfolgt eine Überprüfung, ob angesichts der konkreten Ergebnismenge ein Re-Identifikationsrisiko besteht. Ferner werden dem jeweiligen Schutzbedarf der Daten entsprechende Zugriffsanforderungen an die nutzungsberechtigten Forscher gestellt.

Das FDZ hat, selbst unter Hinzuziehung weiterer Informationen wie Namen oder Versicherungsnummern, technisch keine Möglichkeit, von dem permanenten Pseudonym oder der Arbeitsnummer auf die Identität des Versicherten zu schließen. Dadurch wird bei den im FDZ erfolgenden Datenverarbeitungsvorgängen ein Ansatz verfolgt, der die Beeinträchtigung der Rechte und Freiheiten der betroffenen Personen auf ein Mindestmaß reduziert. Eine gänzlich anonyme Gestaltung der Verarbeitungsvorgänge würde die zur Erreichung der Forschungszwecke notwendige Analyse von periodenübergreifenden Krankengeschichten der Versicherten verhindern und kann daher kein gleich geeignetes Mittel darstellen. Gleiches gilt für die Reduzierung des Datensatzes, beispielsweise im Hinblick auf die Postleitzahl. In dem Fall könnten Forschungsfragen, die an den Wohnort des Versicherten anknüpfen, nicht mehr sinnvoll durchgeführt werden. Eine geringfügige Verfälschung des Datensatzes stellte ebenfalls kein gleich geeignetes Mittel dar. Sie führte zu einem hohen Risiko der Verfälschung von Forschungsergebnissen, welches ferner zunimmt, je kleiner die von einer Forschungsanfrage adressierte Referenzgruppe ist.

Schließlich trägt das BfArM der Voraussetzung der Erforderlichkeit auch in zeitlicher Hinsicht Rechnung, indem die Daten der gesetzlichen Vorschrift des § 303d Abs. 3 SGB V entsprechend spätestens nach 30 Jahren gelöscht werden. Freilich wird ein Löschkonzept implementiert werden, mit dessen Hilfe im jeweiligen Einzelfall geprüft werden kann, ob die gespeicherten DaTraV-Daten für die oben beschriebenen Zwecke noch benötigt werden, oder eine (frühere) Löschung vorzunehmen ist.

Auch soweit personenbezogene Daten der nutzungsberechtigten Forscher verarbeitet werden, ist dies unter jedem der oben genannten Gesichtspunkte erforderlich. Dies gilt insbesondere vor dem Hintergrund, dass zwar teilweise Daten mit Bezug zu natürlichen Personen verarbeitet werden, sich diese aber primär auf die antragsberechtigte Stelle beziehen. Im Rahmen des öffentlichen Antragsregisters wird maximal der Name des Leiters des jeweiligen Forschungsprojekts veröffentlicht. Lediglich die Institution zu publizieren wäre, insbesondere angesichts der hohen Sensibilität der DaTraV-Daten, aus Transparenzgründen nicht gleich geeignet.

5. Verhältnismäßigkeit im engeren Sinne

Die durch die Datenverarbeitungen erfolgenden Beeinträchtigungen des Rechts auf Schutz personenbezogener Daten der betroffenen Personen stehen nicht außer Verhältnis zur Erreichung der im Allgemeininteresse liegenden Zwecke.

Die Verarbeitung der DaTraV-Daten führt dazu, dass zum Zwecke der Forschung im Gesundheitsbereich eine repräsentative Datenbasis bereitgestellt werden kann. Dadurch kann in Zukunft die allgemeine Gesundheit der Bevölkerung geschützt und verbessert werden. Einen besonderen Forschungsvorteil bietet die Möglichkeit des sog. Verschneidens der DaTraV-Daten

mit externen Datensätzen der nutzungsberechtigten Forscher. Eine solche Verschneidung ist unter Einhaltung der datenschutz- und datensicherheitsrechtlichen Anforderungen in einer Form möglich, die eine Herausgabe von Einzeldatensätzen an die Forscher vermeidet und zugleich keine verwertbaren fremden personenbezogenen Daten in den FDZ-Datenbestand aufnimmt.

Die Beeinträchtigung der Rechte und Freiheiten der gesetzlichen Versicherten als betroffene Personen wiegt *prima vista* schwer, da es sich bei den verarbeiteten Daten um hoch sensible Gesundheitsdaten handelt. Demgegenüber stehen indes überragend wichtige Rechtsgüter der Allgemeinheit.

Soweit die Forschung nicht nur der Allgemeinheit, sondern auch dem einzelnen Versicherten, beispielsweise in Gestalt der Vermeidung von Fehlmedikationen und gefährlichen Wechselwirkungen von Arzneimitteln dient, sind die Interessen der gesetzlich Versicherten, des Gesetzgebers und des FDZ gleichgerichtet. Insoweit ist davon auszugehen, dass die Datenverarbeitungen jedenfalls zum Teil auch den betroffenen Personen selbst dienen, was für die Beeinträchtigung ihrer Rechte und Freiheiten einen nicht unerheblichen Ausgleich schafft.

Für die Verhältnismäßigkeit im engeren Sinne spricht ferner, dass es sich bei dem DaTraV-Datenkranz ausschließlich um Daten handelt, die der jeweiligen Krankenkasse des Versicherten ohnehin – und zwar in nicht pseudonymisierter Form – aus dem Datenaustausch mit den Leistungserbringern zum Zwecke der Abrechnung vorliegen. Eine zusätzliche Erhebung oder Verknüpfung von an anderer Stelle gespeicherten Daten mit den personenbezogenen Daten der betroffenen Personen findet nicht statt. Gleiches gilt auch für die Daten der Leistungserbringer als betroffenen Personen, die sie den Krankenkassen ohnehin zu Abrechnungszwecken übermitteln. Mit Blick auf die Leistungserbringer gilt ferner, dass deren Daten nicht sensibel sind und lediglich in einem beruflichen Kontext stehen.

Gegen die Verhältnismäßigkeit im engeren Sinne könnten die mit den technischen Abläufen zwangsläufig verbundenen – im Einzelnen auch teils sehr abstrakten – Risiken sprechen. Die unter Punkt A.VII. vorgenommenen Risikobewertung zeigt indes, dass diesen Risiken mithilfe von Schutz- und Abhilfemaßnahmen adäquat begegnet werden kann.

Insbesondere ist die Verarbeitung der DaTraV-Daten in der Systemarchitektur des FDZ technisch so konzipiert, dass sich die Verarbeitungen personenbezogener Daten durch Verwendung von permanenten Pseudonymen auf ein minimales Maß beschränkt. Keinem der beteiligten Akteure ist es technisch möglich, von den beim FDZ gespeicherten Einzeldatensätzen anhand von Angaben des Versicherten (etwa des Namens oder der Versichertennummer) auf seine Identität zu schließen. Ein Rückschluss auf die Identität des Versicherten käme aus technischer Perspektive nur dann in Betracht, wenn jemand den geheimen Master-Schlüssel, der nur dem

RKI als Vertrauensstelle bekannt ist, erfährt und dann auch in der Lage ist, ihn zum Zwecke der Entschlüsselung anzuwenden.

Mithilfe von technischen Maßnahmen und durch im jeweiligen Einzelfall stattfindende Prüfungen der Ergebnismengen durch FDZ-Analysten wird sichergestellt, dass das Re-Identifikationsrisiko einzelner Versicherten weitestgehend ausgeschlossen wird. Im Übrigen sorgen gesetzliche Vorschriften als zusätzliches Korrektiv dafür, dass die Auswirkungen einer theoretischen Re-Identifikation für die betroffene Person möglichst gering bleiben. Beispielsweise muss das FDZ sicherstellen, dass DaTraV-Daten nur Nutzungsberechtigten zugänglich gemacht werden, die einer Geheimhaltungspflicht nach § 203 des Strafgesetzbuches unterliegen oder vor dem Zugang gemäß § 1 Abs. 2, 3 und 4 Nr. 2 des Verpflichtungsgesetzes zur Geheimhaltung verpflichtet wurden.

Auch mit Blick auf die „Standardrisiken", beispielsweise das Risiko eines Hacker-Angriffs, ergibt sich keine gravierende zusätzliche Gefahr durch die Erhebung und Speicherung der Daten beim FDZ. Denn die Gefahr eines Hacker-Angriffs besteht auch schon hinsichtlich des IT-Systems der jeweiligen Krankenkasse, auf dem dieselben Daten in nicht pseudonymisierter Form gespeichert sind. Ein Angriff auf die Leistungserbringer wird sich ebenfalls vorrangig anbieten, da sie hinsichtlich der Sicherheitsmechanismen in ihrer IT-Infrastruktur das schwächste Glied in der Kette derjenigen bilden dürften, welche über diese Art von Daten verfügen.

Auch hinsichtlich der personenbezogenen Daten der nutzungsberechtigten Forscher sind die beim FDZ erfolgenden Verarbeitungen verhältnismäßig im engeren Sinne. Zum einen stellen sie die Anträge freiwillig, was auch mit Blick auf dabei verarbeiteten personenbezogenen Daten Ausdruck ihres Rechts auf informationelle Selbstbestimmung ist. Zum anderen sind deren Daten wenig sensibel und die Beeinträchtigung ihrer Rechte und Freiheiten damit insgesamt als gering zu bewerten.

Aus alldem folgt, dass die Datenverarbeitungen und die mit ihnen verbundenen Eingriffe in die Rechte der betroffenen Personen im Fall des FDZ als verhältnismäßig zu bewerten sind.

V. Datenschutzrechtliche Bewertung (kein Pflichtbestandteil des DSFA-Berichts gemäß Art. 35 Abs. 7 DS-GVO)

1. Rechtsgrundlagen für die einzelnen Verarbeitungen personenbezogener Daten

Rechtsgrundlage für die Verarbeitung der DaTraV-Daten sind die §§ 303a ff. SGB V in Verbindung mit der DaTraV, die das Verfahren der Datentransparenz abschließend regeln – mit Blick auf das hier im Fokus stehende FDZ insbesondere § 303d, § 303e SGB V und § 4 ff.

DaTraV. Es handelt sich um eine Rechtsgrundlage im Sinne des Artikel 6 Abs. 1 lit. e DSGVO und hinsichtlich der Verarbeitungen der gesundheitsbezogenen Daten im Sinne des Art. 9 Abs. 2 lit. i und j, Abs. 4 DSGVO i.V.m. § 22 Absatz 1 Nummer 1 lit. c BDSG.

2. Gewährleistung der Betroffenenrechte

Die Rechte der betroffenen Personen sowie die entsprechenden Pflichten des datenschutzrechtlich Verantwortlichen bestimmen sich nach Kapitel 3 der DS-GVO. Diese sind insbesondere das Recht auf Auskunft (Art. 15 DS-GVO), Berichtigung (Art. 16 DS-GVO), Löschung (Art. 17 DS-GVO), Einschränkung der Verarbeitung (Art. 18 DS-GVO) sowie Datenübertragbarkeit (Art. 20 DS-GVO). Grundsätzlich hat jeder datenschutzrechtlich Verantwortliche diese Rechte der betroffenen Personen zu erfüllen. Eine Ausnahme hiervon regelt Art. 11 DS-GVO. Dessen Abs. 1 bestimmt:

> „Ist für die Zwecke, für die ein Verantwortlicher personenbezogene Daten verarbeitet, die Identifizierung der betroffenen Person durch den Verantwortlichen nicht oder nicht mehr erforderlich, so ist dieser nicht verpflichtet, zur bloßen Einhaltung dieser Verordnung zusätzliche Informationen aufzubewahren, einzuholen oder zu verarbeiten, um die betroffene Person zu identifizieren."

Kann der Verantwortliche in diesen Fällen nachweisen, dass er nicht in der Lage ist, die betroffene Person zu identifizieren, unterrichtet er sie gemäß Art. 11 Abs. 2 Satz 1 DS-GVO hierüber. Gemäß Art. 11 Abs. 2 Satz 2 DS-GVO finden insofern die Artikel 15 bis 20 keine Anwendung, es sei denn, die betroffene Person stellt zur Ausübung ihrer in diesen Artikeln niedergelegten Rechte zusätzliche Informationen bereit, die ihre Identifizierung ermöglichen. D.h. der Verantwortliche muss die o.A. Betroffenenrechte nicht erfüllen, wenn er nachweisen kann, dass er eine betroffene Person nicht identifizieren kann, es sei denn, dass diese selbst die notwendigen Informationen zur Identifizierung beibringt.

Das BfArM muss die Betroffenenrechte der betroffenen Personen nicht erfüllen. Denn es ist aufgrund des im RKI erstellten permanenten Pseudonyms nicht in der Lage, die betroffene Person – namentlich den gesetzlich Versicherten – zu identifizieren. Die betroffene Person kann vorliegend auch nicht selbst die zusätzlichen Informationen bereitstellen, mit denen sie eindeutig identifizierbar wird: Die für die Gewährleistung eines Betroffenenrechts notwendige zusätzliche Information kann vorliegend nicht in der Versichertennummer bestehen, da auch anhand dieser für das BfArM eine Identifikation des Betroffenen nicht möglich ist. Eine Identifizierung wäre nur möglich, wenn die betroffene Person dem BfArM das permanente Pseudonym mitteilen und nachweisen könnte, dass es zu ihrem Einzeldatensatz gehört.

Hypothetisch wäre es denkbar, dass die betroffene Person dieses permanente Pseudonym aufgrund eines Auskunftsanspruchs gegenüber dem RKI erfahren könnte. Dies ist indes unmöglich: Sowohl die Versichertennummer als auch die Namen der Versicherten werden beim RKI nur sehr kurzzeitig und nur innerhalb der Pseudonymisierungsoftware entschlüsselt und nicht gespeichert. Betroffene Personen können somit auch beim RKI anhand ihrer Versichertennummer oder ihres Namens nicht identifiziert werden, sodass auch das RKI aus faktischen Gründen keine Auskunft zum periodenübergreifenden Pseudonym geben kann. Gleiches gilt für den Spitzenverband Bund der Krankenkassen, dem die Versichertennummern seitens der Krankenkassen nicht übermittelt werden. Dieser verfügt auch nicht über den entsprechenden Schlüssel, um die Versichertennummer aus dem Lieferpseudonym zu entschlüsseln. Dieser verbleibt bei der jeweiligen Krankenkasse, die ihn allerdings unmittelbar nach Übermittlung der Daten an den Spitzenverband Bund der Krankenkassen löscht. Daraus folgt, dass die betroffene Person auch aufgrund von Auskunftsansprüchen gegen die anderen beteiligten Akteure keine Informationen erlangen kann, anhand derer ihre Identifikation dem BfArM möglich wäre.

Eine eindeutige Identifikation ist auch nicht aufgrund anderer, seitens der betroffenen Person zur Verfügung gestellter Informationen möglich. Vorstellbar ist, dass eine betroffene Person gewisse Parameter (Alter, Geschlecht, Postleitzahl, Erkrankungen, Daten von Krankenhausaufenthalten etc.) preisgibt, die nur auf eine einzige Person im IT-System des FDZ zutreffen. Diese Angaben können aber ggf. doch auf mehrere Personen zutreffen oder missbräuchlich von dem Antragsteller verwendet werden. Dabei kann es sich beispielsweise um Informationen seines Nachbaren handeln, von denen er zufällig Kenntnis erlangt hat. Jedenfalls kann das BfArM anhand dieser Parameter keine *zuverlässige* Identifizierung der betroffenen Person vornehmen, sodass keine eindeutige Authentifizierung dieser Person möglich ist. Unschärfen in der Zuordnung braucht der Verantwortliche regelmäßig nicht hinzunehmen, sodass das BfArM die Betroffenenrechte gemäß Art. 11 DS-GVO nicht erfüllen muss.

Dies gilt bei den Datenverarbeitungsvorgängen des FDZ umso mehr, da die Betroffenenrechte unproblematisch gegenüber der jeweiligen Krankenkasse des gesetzlich Versicherten geltend gemacht werden können. Das FDZ verarbeitet keine Daten, die über den bei den Krankenkassen verarbeiteten Datenkranz hinausgingen, sodass den Betroffenenrechten durch die Geltendmachung gegenüber den Krankenkassen hinreichend Rechnung getragen wird.

VI. Standpunkt der betroffenen Personen

Gemäß Art. 35 Abs. 9 DS-GVO kann der Verantwortliche den Standpunkt der betroffenen Personen oder ihrer Vertreter zu der beabsichtigten Verarbeitung einholen, um deren Sichtweise in Erfahrung zu bringen und mögliche Kritikpunkte frühzeitig zu identifizieren und Abhilfe schaffen zu können. Die Einholung der Standpunkte betroffener Personen bzw. ihrer Vertreter

ist im Zuge des Gesetzgebungsverfahrens des Digitale-Versorgung-Gesetzes erfolgt, mit dessen Inkrafttreten das Verfahren der Datentransparenz geändert wurde. Auch der BfDI hat sich in einer Stellungnahme zum Gesetzesentwurf geäußert. Eine darüberhinausgehende Anhörung betroffener Personen oder ihrer Vertreter durch das BfArM als Verantwortlicher der Datenverarbeitungen war angesichts der gesetzlichen Verpflichtung zur Verarbeitung des festgelegten Datenkranzes in §§ 303a ff. SGB V nicht angezeigt.

VII. Bewertung der Risiken für die Rechte und Freiheiten der betroffenen Personen, Art. 35 Abs. 7 lit. c DS-GVO

1. Angewandte Methode

Bei der Durchführung der DSFA wurde das Standard-Datenschutzmodell (SDM) in der Version 2.0b zugrunde gelegt. Das SDM verdichtet und systematisiert sämtliche Anforderungen der DS-GVO in Form von sieben Gewährleistungszielen – Datenminimierung, Verfügbarkeit, Integrität, Vertraulichkeit, Nichtverkettung, Transparenz und Intervenierbarkeit. Im Rahmen des SDM werden jedem Gewährleistungsziel spezifische technische und organisatorische Abhilfemaßnahmen zugeordnet, mittels derer das Ziel und die dahinterstehenden Anforderungen der DS-GVO gewährleistet und der Eintritt von Schadensereignissen verhindert werden kann. Ergänzend wurde die die vom BSI entwickelten Standards für den IT-Grundschutz herangezogen.

2. Risikoidentifikation, -Analyse und -Bewertung[7]

Für die Risikoidentifikation, für die Risikoanalyse und für die Risikobewertung ist auf die in der **Anlage 1** beigefügte Excel-Tabelle zu verweisen.

VIII. Zur Bewältigung der Risiken geplante Abhilfemaßnahmen, Art. 35 Abs. 7 lit. d DS-GVO[8]

1. Maßnahmen zur Behandlung von fundamentalen Risikogruppen

Wie aus der detaillierten Risikoanalyse (Anlage [1]) hervorgeht, gibt es eine Reihe von Risikobereichen, deren Einzelrisiken ähnlich sind. Beispielsweise unterscheiden sich die Risiken

---

7 Die nachstehenden Ausführungen stammen von Dr. Oliver Stiemerling, ecambria systems.

8 Die nachstehenden Ausführungen stammen von Dr. Oliver Stiemerling, ecambria systems.

durch böswillige interne Akteure nur durch die jeweils unterschiedlichen Zugriffsrechte (Administrator, FDZ-Analyst). Folgende Risikobereiche sind näher zu betrachten:

- Böswillige Handlungen interner Akteure
- Böswillige Handlungen externer Dienstleister
- Böswillige Handlungen krimineller Angreifer
- Böswillige Handlungen fremder staatlicher Angreifer
- Böswillige Handlungen antragsberechtigter Stellen
- Aktuelle und zukünftige „Begehrlichkeiten" eigener staatlicher Akteure
- Risiken durch die Erfüllung von Betroffenenrechten, insb. Auskunft
- Fehler interner Akteure und externer Dienstleister
- Risiken durch Dritte (Krankenkassen, Spitzenverband, RKI), insb. De-Pseudonymisierung
- Risiken durch technischen Fortschritt in den nächsten 30 Jahren, insb. Quantencomputer

Im Folgenden werden für die jeweiligen Risikobereiche die geplanten Einzelmaßnahmen bzw. auch Maßnahmenbündel genannt.

Ziel der Abhilfemaßnahmen ist es, Risiken mit einer kritischen Einstufung (rot) zumindest auf Risiken mittleren Grades (gelb) im Sinne des SDM (s.o.) zu senken.

2. Grundsätzliche Feststellung zur Sicherheit durch die gesetzlich vorgeschriebene Pseudonymisierung und organisatorische Trennung

Die im Gesetz festgelegte Pflicht zur Pseudonymisierung und organisatorischen Trennung (RKI als Vertrauensstelle, getrennte Lieferungen von Pseudonymen und Inhaltsdaten) stellt bereits eine sinnvolle Maßnahme dar, um das Risiko eines Bruchs der Vertraulichkeit in quasi allen Risikobereichen zu vermindern. Durch die Pseudonymisierung wird es einem böswilligen Akteur mit legalem oder auch kriminellem Zugriff auf den FDZ-Datenbestand wesentlich erschwert, die dort gespeicherten Daten zu eigenen oder fremden Zwecken zu missbrauchen. Eine einfache Abfrage im Sinne „Zeige mir alle Daten von Otto Müller, wohnhaft Peterstr. 12, 50932 Köln" ist durch die Pseudonymisierung auf dem FDZ-Datenbestand unmöglich.

Allerdings haben statistische Untersuchungen des FDZ ergeben, dass mit bereits geringem Zusatzwissen (Postleitzahl, Geschlecht, Alter, eine Kombination von zwei Diagnosen) eine De-Pseudonymisierung einzelner Betroffener regelmäßig möglich ist. Ein böswilliger Akteur mit Zugriff auf den FDZ-Datenbestand könnte also mit diesem Zusatzwissen die volle Krankheitshistorie einer konkreten, betroffenen Person abfragen.

Dieses Risiko ist für die hiesige Untersuchung relevant, da Alter, PLZ, Geschlecht und zwei Krankheiten mit 10% und 5% Auftretenswahrscheinlichkeit ausreichen, um im weit überwiegenden Teil der PLZ-Gebiete eine Person eindeutig zu identifizieren. Das ergibt sich aus der nachstehenden Prüfung:

	Faktor		20000 Personen im PLZ-Bereich (Mittelwert < 10.000, Median < 7.000)
Alter	70		
Geschlecht	2		
Diagnose 1	20		Eine Krankheit, die jeder 20. Bürger hat
Diagnose 2	10		Eine Krankheit, die jeder 10. Bürger hat
Gesamtfaktor	28000		0,71428571 Personen, die alle Merkmale erfüllen

Abbildung 1: Plausibilisierung Angaben FDZ zur De-Pseudonymisierung auf dem Datenbestand

So würde es beispielsweise einem Administrator mit Lesezugriff auf die Datenbank leichtfallen, einen konkreten Nachbarn oder Verwandten im Datenbestand zu identifizieren und die gesamte Krankenhistorie abzufragen. Dieses Risiko steigt, je weiter der FDZ-Datenbestand in die Vergangenheit zurückreicht, da dann mehr Details der Krankheitshistorie verfügbar sind, derzeit ein Datenbestand von bis zu 30 Jahren.

Unabhängig vom Identifikationsmittel der Diagnose können auch weitere sensible Informationen – etwa Krankenhaus- und Arztbesuche oder andere im Datenbestand enthaltene differenzierende Details – die Identifikation von Einzelpersonen ermöglichen. Die Konsequenz dieses dem Datenbestand immanenten De-Pseudonymisierungsrisikos ist, dass

a) unberechtigte Zugriffe von böswilligen Dritten und

b) missbräuchliche, berechtigte Zugriffe durch böswillige Innentäter oder Dienstleister

unbedingt verhindert werden müssen. Den konkreten Schutzmaßnahmen mit dem Ziel der Wahrung der Vertraulichkeit (und auch Nicht-Verkettbarkeit) kommt somit eine wesentliche Bedeutung zu.

3. Grundsätzliche Feststellungen zu den verschiedenen Gewährleistungszielen

Aus Sicht der Betroffenen stellen die Vertraulichkeit und die Nicht-Verkettbarkeit wichtige Schutzziele dar, weil eine Offenlegung hochsensibler Gesundheitsdaten existentielle Nachteile zur Folge haben kann. Die Verletzung der Integrität und der Verfügbarkeit der Daten hat hingegen keine negativen Konsequenzen für den Einzelnen, da die Daten an sich für die Betroffenen selbst keinerlei operative Bedeutung haben. Auch das Gewährleistungsziel Datenminimierung wird hier nicht weiter betrachtet, da der Umfang der Datenhaltung in Bezug auf die gespeicherten Attribute und Verknüpfungen gesetzlich vorgegeben ist und nicht durch technische oder organisatorische Maßnahmen eingeschränkt werden kann.

Schließlich ist das Ziel Transparenz aus technischer Sicht alleine auf Basis des FDZ-Datenbestandes nicht sachgerecht zu gewährleisten, da ein Antragsteller lediglich eine Identifizierung anhand von Inhaltsdaten (Postleitzahl, Geschlecht, Alter, Diagnosen) verlangen könnte. Auf dieser Basis ist eine sichere Authentifizierung vor der Herausgabe des vollständigen Datensatzes nicht möglich, da die Daten auch im Besitz eines unberechtigten Dritten sein können.

Ein Antragsteller würde beispielsweise mitteilen: „Ich bin der 76 Jahre alte Mann aus 50932 Köln, der eine künstliche Hüfte und Bluthochdruck hat – bitte geben Sie mir alle meine weiteren Gesundheitsdaten". Das FDZ hätte in einem solchen Fall keine Möglichkeit zu überprüfen, ob es sich beim Antragsteller tatsächlich um dieselbe natürliche Person handelt oder um einen Dritten, der über diese Information (unrechtmäßig) verfügt. In solchen Fällen kann das FDZ mangels Kenntnis des konkreten Namens keine Überprüfung anhand von Ausweisdokumenten o.ä. vornehmen.

Auch wenn sich ein Antragsteller alternativ mit einem über die Krankenkassen bzw. das RKI ermittelten Pseudonym (konkret: dem periodenübergreifenden Pseudonym) identifiziert, kann das FDZ den Antragsteller alleine auf Basis des eigenen Datenbestands nicht ausreichend sicher authentifizieren. Es müsste den Krankenkassen und dem RKI vertrauen, dass diese eine für den Schutzbedarf angemessene Authentifizierung vorgenommen haben. Ebenfalls nicht bewerten könnte das FDZ anhand seines eigenen Datenbestands, ob der Antragsteller sich das entsprechende Pseudonym nicht illegal bei der tatsächlichen natürlichen Person beschafft hat, die das Pseudonym vorher von Krankenkasse/RKI erfragt hat. Zudem sollte das Pseudonymisierungsverfahren bei den Krankenkassen und dem RKI idealerweise so konfiguriert werden, dass eine spätere Rekonstruktion von Lieferpseudonym oder jahresübergreifendem Pseudonym aus einer von einem potentiellen Antragsteller übermittelten Versichertennummer o.ä. nicht möglich ist.

4. Grundsätzliche Feststellungen zum Risiko für die verschiedenen Klassen von Betroffenen

Bei der Datenverarbeitung im FDZ stehen insbesondere die Risiken der gesetzlich Versicherten im Vordergrund, da ein Bruch der Vertraulichkeit bzw. Nicht-Verkettbarkeit der Daten existentiell negative Konsequenzen haben kann.

Die Risiken für die ebenfalls im Datenbestand pseudonym gespeicherten Leistungserbringer (z.B. Ärzte) erscheinen aus technischer Sicht wesentlich geringer, da die in ihrer Rolle als Leistungserbringer gespeicherten Daten keine Gesundheitsdaten darstellen, sondern lediglich Schlussfolgerungen über Quantität und Qualität ihrer beruflichen Tätigkeit erlauben. Eine Analyse dieser Daten kann indes – beispielsweise mit dem Ziel der Leistungskontrolle oder Betrugserkennung – erhebliche Auswirkungen auf den Betroffenen haben. Allerdings werden dieselben Daten unpseudonymisiert beim Spitzenverband Bund der Krankenkassen verarbeitet, so

dass die Verarbeitung im FDZ prinzipiell keine Erhöhung des bereits bestehenden Risikos verursacht und damit keine Maßnahme des FDZ das Risiko vermindern könnte.

5. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]

- [redacted]

  [redacted]

  [redacted]

- [redacted]

  [redacted]

  - [redacted]

  - [redacted]

  - [redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

IX. Nachhaltige Sicherung des Datenschutzes und Evaluation

Das Datenschutzkonzept sieht kontinuierliche Überprüfungen/Evaluationen der getroffenen Maßnahmen und der DSFA vor.

In regelmäßigen Abständen müssen Kernelemente des Datenschutzes und der Datensicherheit im Rahmen eines effektiven Datenschutzmanagements überprüft werden. Insbesondere sind die Abhilfemaßnahmen, welche die im Rahmen dieser DSFA identifizierten Risiken minimieren, anhand des jeweiligen Stands der Technik fortlaufend zu überprüfen und ggf. anzupassen. Dies gilt im Speziellen mit Blick auf Technologien, die die Sicherheit der derzeitigen Pseudonymisierungsverfahren in Frage stellen könnten.

[redacted]