LABIMI/F

Deliverable

Johannes Drepper [a]
unter Mitarbeit von
Frank Dickmann[b], Philipp Weil[b]
[a] TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.
[b] Institut für Medizinische Informatik, Universitätsmedizin Göttingen

D5.1:Datenschutzkonzept[1]

Titel	D5.1: Datenschutzkonzept
Autor(en)	Johannes Drepper; Frank Dickmann
Editor(en)	Philipp Weil, Romanus Grütz
Datum	20.09.2013

A: Status des Dokuments

Version 1.0

B: Bezug zum Projektplan

M5.1: Datenschutzkonzept

[1]Dieses Dokument wurde im Rahmen des Projekts LABIMI/F erstellt. Das Projekt LABIMI/F wird gefördert von der Deutschen Forschungsgemeinschaft (DFG) unter dem Förderkennzeichen RI1000/2-1.

C: Abstract

Das vorliegende Dokument beschreibt ein generisches Datenschutzkonzept für das Anwendungsfeld der Langzeitarchivierung biomedizinischer Forschungsdaten. Als beispielhafte Anwendungsfälle und daraus abgeleitete Zweckbestimmungen werden die Erhebung und Nutzung von Bild- und Genomdaten im biomedizinischen Kontext betrachtet. Eine ausführliche Analyse der rechtlichen Rahmenbedingungen erlaubt die Ableitung rechtskonformer Konzepte für eine Vielzahl unterschiedlicher Forschungsprojekte. Für die hier beispielhaft betrachteten Anwendungsfälle werden vor dem Hintergrund der hierfür relevanten rechtlichen Rahmenbedingungen drei verschiedene rechtskonforme Modelle einer technischen und organisatorischen Umsetzung dargestellt und diskutiert.

D: Änderungen

Version	Datum	Name	Kurzbeschreibung
0.0.1	05.03.2013	Frank Dickmann	Struktur
0.0.2	27.06.2013	Johannes Drepper	Erste Version
0.2	05.08.2013	Johannes Drepper	Überarbeitung und Ergänzung nach Kommentierung durch Frank Dickmann, Philipp Weil und Romanus Grütz
1.0	20.09.2013	Johannes Drepper	Überarbeitung nach Kommentierung durch Philipp Weil und Romanus Grütz und Ergänzung eines Abstracts

E: Inhaltsverzeichnis

1 Einleitung

Digitale Forschungsdaten werden in der Biomedizin wie in nahezu allen wissenschaftlichen Disziplinen in exponentiell steigendem Ausmaß erzeugt. Deren nachhaltige Nutzbarmachung ist für die Forschung gegenwärtig ein Thema von besonderer Bedeutung. Dabei ist zum einen schon die Gewährleistung einer sicheren Archivierung gemäß den entsprechenden rechtlichen wie wissenschaftsinternen Vorgaben und angesichts der zunehmenden Datenmengen eine besondere Herausforderung. Zum anderen geraten aber auch immer mehr Anwendungsfälle in den Blick, in denen die Daten für neue wissenschaftliche Vorhaben innerhalb der Forschergemeinschaft nachgenutzt werden können. Dies ist insbesondere hinsichtlich des Potentials genetischer Daten und der Entwicklung hin zu einer individualisierten Medizin nachvollziehbar. Aber auch andere Datensammlungen der Biomedizin, wie z. B. Bilddaten, bergen erhebliches Nachnutzungspotential. Dabei müssen aus diesen Daten nicht immer direkt neue Erkenntnisse gewonnen werden, häufig können sie auch bei der Hypothesenbildung oder der Rekrutierung passender Probanden für neue Studien helfen.

Infrastrukturen, die alle diese Anwendungsfälle unterstützen, unterliegen einer Vielzahl rechtlicher und technischer Anforderungen. Neben den urheber- und verwertungsrechtlichen Aspekten sind insbesondere datenschutzrechtliche Anforderungen zu berücksichtigen, da die für die Nachnutzung vorzuhaltenden Daten in der biomedizinischen Forschung häufig als besonders schützenswerte Daten gemäß § 3 (9) BDSG anzusehen sind. Als technische Anforderungen können daraus schon Vorgaben zur Anonymisierung oder Pseudonymisierung abgeleitet werden (vergl. § 3a BDSG). Darüber hinaus ist eine Nachnutzung aber nur dann möglich, wenn z. B. aussagekräftige und auch technisch auswertbare Metadaten vorliegen. Hinzu kommen technische Anforderungen hinsichtlich der Speicher- und Verarbeitungskapazität wie auch der nötigen Anbindung an öffentliche Netzwerke und ggf. auch der Umsetzung von Zugriffsbeschränkungen. Aus den genannten rechtlichen wie technischen Anforderungen lassen sich wiederum organisatorische und wirtschaftliche Vorgaben für solche Infrastrukturen ableiten.

In dem von der DFG geförderten Projekt zur Langzeitarchivierung biomedizinischer Forschungsdaten (LABIMI/F) wird den vorgenannten Fragen und Anforderungen am Beispiel zweier Use Cases nachgegangen, die zum einen die Nachnutzung genetischer Daten und zum anderen die Nachnutzung von Bilddaten zum Gegenstand haben. Im Ergebnis soll u. a. ein Betriebskonzept erarbeitet werden, welches auf einen Großteil dieser Fragestellungen eingeht. Dieses Datenschutzkonzept und ein ergänzendes Service-Level-Agreement (SLA) sind notwendigerweise in enger Abstimmung mit dem Betriebskonzept entstanden.

Das Projekt LABIMI/F wird geleitet von Prof. Rienhoff vom Institut für Medizinische Informatik der Universitätsmedizin Göttingen. Weitere Partner sind die Universitätskliniken Magdeburg und Schleswig-Holstein (Standort Kiel) sowie die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), das Zuse-Institut Berlin und die TMF.

Das vorliegende Dokument untersucht die datenschutzrechtlichen Rahmenbedingungen für den Betrieb einer Langzeitarchivierungs-Infrastruktur (LZA-Infrastruktur) und skizziert dar-

über hinaus auch, welche Festlegungen in einem konkreten Datenschutzkonzept für eine spätere Abstimmung mit den zuständigen Aufsichtsbehörden zu treffen sind.

2 Anwendungsfälle und Zweckbestimmung

2.1 *Use Case Biomedizinische Bilddaten*

Die folgende Darstellung stützt sich auf den im Rahmen des Projekts am 29.2.2012 in Magdeburg durchgeführten Workshop zum Forschungsdatenmanagement in der medizinischen Bildverarbeitung sowie das zugehörige Deliverable[2].

In der medizinischen Bildgebung wird eine Vielzahl unterschiedlicher technischer Bildgebungsverfahren genutzt, die sich hinsichtlich der damit verbundenen Risiken für die Probanden unterscheiden lassen in invasive Verfahren, die lediglich im klinischen Kontext zulässig sind, und nicht-invasive Verfahren, die auch in der nicht-klinischen Forschung eingesetzt werden können. Zur ersten Gruppe gehören z. B. Verfahren wie Röntgen, Computertomographie (CT) oder die Positronenemissionstomographie (PET), die die Probanden einer gewissen radioaktiven Strahlenbelastung aussetzen. Ein häufig genutztes Verfahren der zweiten Gruppe ist die Magnetresonanztomographie (MRT), die ohne eine solche Strahlenbelastung auskommt. Im vorliegenden Projekt wurde vornehmlich der Anwendungsfall der Langzeitarchivierung nicht-klinisch gewonnener Bilddaten untersucht, was zu einer entsprechenden Einschränkung der berücksichtigten Bildgebungsverfahren führt.

Für eine Langzeitarchivierung von Forschungsdaten müssen neben technischen Lösungen vor allem Metadatenmodelle erfasst werden, welche es ermöglichen, Forschungsdaten sinnvoll in eine Datenbasis einzuordnen und mit Hilfe dedizierter Suchparameter zu einem beliebigen Zeitpunkt wiederaufzufinden. Dadurch soll den Forschungseinrichtungen die institutionsübergreifende Wiederverwendung und Nachnutzung bereits erzeugter Daten ermöglicht und Transparenz bezüglich des Entstehungsprozesses digitaler wissenschaftlicher Resultate gewährleistet werden. Aus der Perspektive des Datenschutzes ist zu berücksichtigen, dass sowohl die Bildinformationen selbst als auch die Metadaten hinsichtlich des Reidentifikationspotentials zu berücksichtigen sind.

Zu den Metadaten können auch umfassende Daten über das exakte experimentelle Setting einer Untersuchung gehören. So ist in einer Studie mit funktioneller Bildgebung (fMRI) beispielsweise eine Beschreibung der verschiedenen Aufnahmebedingungen für eine Differenzberechnung der Indikatorparameter der Hirnaktivität notwendig. In solchen experimentellen Bedingungen ist häufig eine sehr genaue Synchronisation aller Abläufe und Messungen nötig, so dass auch exakte Zeitangaben zu den relevanten Metadaten gehören und nur in Kenntnis der exakten Versuchsbedingungen ggf. in angepasster Weise vergröbert werden können.

Eine Nachnutzung umfassend dokumentierter Bilddaten erscheint sowohl innerhalb einer Forschungseinrichtung lohnenswert als auch einrichtungsübergreifend oder in einem internationalen Kontext, was jeweils zu spezifischen datenschutzrechtlichen Anforderungen führt.

[2] Alle Dokumente zum Workshop und die Deliverables siehe http://informatik.med.uni-magdeburg.de/lza.html

Dabei können mit dem Aufbau einer übergreifenden LZA-Infrastruktur zwei ganz unterschiedliche Ziele verfolgt werden, die auch zu unterschiedlichen datenschutzrechtlichen Anforderungen führen. Zum einen können Forscher in Bezug auf die vorliegenden Archivierungspflichten gesetzlicher oder auch untergesetzlicher Art (siehe detaillierte Erörterung weiter unten) unterstützt werden. Eine solche Archivierung ist beispielsweise notwendig, um später nachweisen zu können, dass korrekt geforscht und publiziert wurde. Zum anderen besteht aber auch großes Interesse innerhalb der wissenschaftlichen Community an der Nachnutzung solcher Daten, insbesondere dann, wenn sie gut dokumentiert und mit ausführlichen und standardisierten Metadaten versehen sind[3]. Während im ersten Fall nur der Forscher selbst Zugriff auf die Daten haben darf, gilt diese Restriktion für das zweite Anwendungsfeld gerade nicht. Entsprechend sind für die übergreifende Nachnutzung andere und weitergehende datenschutzrechtliche Rahmenbedingungen zu berücksichtigen.

2.2 *Use Case Genomdaten*

Die Darstellung des Use Case der Langzeitarchivierung von im Forschungskontext erhobenen Genomdaten stützt sich auf den im Rahmen des vorliegenden Projekts durchgeführten Workshop am 27.3.2012 zum „Forschungsdatenmanagement biomedizinischer Genomdaten" in Kiel[4].

Auch bezüglich der Archivierung der Sequenzierungsdaten sind die beiden schon für Bilddaten beschriebenen Anwendungsfelder zu unterscheiden: Zum einen die Archivierung zum Zwecke der späteren Nachvollziehbarkeit und zum anderen ein Data Sharing im Sinne einer übergreifenden Nachnutzung durch die wissenschaftliche Community. Auch hier gilt, dass sich unterschiedliche Anforderungen des Datenschutzes und bezüglich des Detailgrads der zu speichernden Daten ergeben. Insbesondere für die Use Cases zur Nachnutzung der Daten durch andere Forscher ist eine Suchmöglichkeit in aussagekräftigen und möglichst standardisierten Metadaten essenziell. Allerdings ist dabei der Aufwand der Dateneingabe für den Forscher in einem sinnvollen Verhältnis zu dem Wert einer späteren Nachnutzung zu halten. Ein beispielhafte Metadatenschema wurde im Rahmen des vorliegenden Projekts entwickelt und wird im Rahmen des Deliverable 5.2 veröffentlicht[5].

Aufgrund der im Verhältnis zur Bildgebung relativ neuen Methodik der Erhebung genetischer Daten besteht auch für das Anwendungsfeld der klassischen Archivierung zwecks Nachvollziehbarkeit noch Unsicherheit hinsichtlich der zu verwendenden Datenformate und der Notwendigkeit der Speicherung von im Analyseprozess entstehenden Zwischenergebnissen. Dies betrifft z. B. die im Rahmen der Sequenzierung anfallenden Bilddaten, die heute z. T. noch parallel zu den eigentlichen Sequenz- und Qualitätsdaten aufbewahrt werden. Damit zusammenhängend bestehen auch noch vergleichsweise wenig Erfahrungen dazu, in welcher Form und mit welchen Hilfsmitteln später einmal ein Nachweis der korrekten Durchführung des heutigen Forschungsprojekts zu führen ist.

[3] Hinweise hierzu finden sich auch in Interviews, veröffentlicht als Deliverable 4.4 auf http://www.labimi-f.med.uni-goettingen.de/documents.html
[4] Agenda, Foliensätze und zusammenfassende Beschreibung siehe http://www.labimi-f.med.uni-goettingen.de/documents.html
[5] siehe http://www.labimi-f.med.uni-goettingen.de/documents.html

Das Anwendungsfeld der einrichtungsübergreifenden Nachnutzung genetischer Daten ist noch relativ jung. Allerdings ist eine sehr dynamische Entwicklung feststellbar. Beispielhaft hierfür wäre die vom NIH entwickelte und bereitgestellte Database of Genotypes and Phenotypes (dbGaP) [1], die GenBank des NCBI, das European Nucleotide Archive (ENA) oder die Protein Data Bank (PDB) zu nennen. Allerdings sind mit der Bereitstellung individueller genetischer Datensätze von Probanden außerhalb der USA noch erhebliche rechtliche Probleme verbunden [2]. Im Rahmen des Workshops wurden zwei beispielhafte Studien vorgestellt, in denen durch die Nachnutzung eigener Daten aus genomweiten Assoziationsstudien (GWAS) eine weitergehende Aufklärung der genetischen Ursachen der Hämochromatose und von Gallensteinleiden erreicht werden konnte.

Als ein wichtiges Kriterium der Nachnutzbarkeit wird neben den notwendigen Metadaten auch die Verfügbarkeit umfangreicher und ebenfalls gut dokumentierter klinischer oder Phänotyp-Daten angesehen, so dass das Vorhandensein solcher Datenbestände samt deren Verlinkung zu den genetischen Daten und die darauf nötigen Zugriffe ebenfalls datenschutzrechtlich zu berücksichtigen sind.

3 Rechtliche Rahmenbedingungen

Aus datenschutzrechtlicher Sicht ist zunächst für die gegebenen Anwendungsfälle und Verarbeitungsvorgänge zu prüfen, ob das Datenschutzrecht zur Anwendung kommen kann. Dies ist dann nicht der Fall, wenn es spezialgesetzliche Regelungen gibt, die bestimmte Verarbeitungsvorgänge festlegen, da das Datenschutzgesetz nur nachgeordnet, also subsidiär gilt (vergl. § 1 (3) BDSG). Im Forschungskontext zu betrachtende Rechtsgrundlagen sind beispielsweise:

- AMG, MPG
- Röntgenverordnung (RöV), Strahlenschutzverordnung (StrlSchV)
- Musterberufsordnung für Ärzte (MBO)
- Gentechnikgesetz, Gendiagnostikgesetz
- Landeskrankenhausgesetze

Für die rechtlichen Grundlagen der Archivierung der Daten aus klinischen Studien sei an dieser Stelle auf das im Auftrag der TMF erstellte Gutachten von Dierks hingewiesen [3]. Die rechtlichen Rahmenbedingungen der Archivierung genetischer Daten wurden von Mathieu et al. zusammengefasst [4]. Wenn die genannten Regelungen für eine Nachnutzung langzeitarchivierter Daten keine Rechtsgrundlage bieten, kommen die Bestimmungen des Datenschutzrechts zur Anwendung. Hier ist in Deutschland zu unterscheiden zwischen privatrechtlich und öffentlich begründeten Einrichtungen. Für erstere wie auch die öffentlichen Einrichtungen des Bundes ist das BDSG anzuwenden, für die öffentlichen Einrichtungen der Länder im Regelfall die Landesdatenschutzgesetze (LDSG). Der Regelungsgehalt der verschiedenen LDSG und des BDSG in Bezug auf die Forschung ist weitgehend identisch, was aber relevante Detailunterschiede leider nicht ausschließt. Im Weiteren wird, von besonderen Ausnahmen abgesehen, stellvertretend nur auf das BDSG Bezug genommen.

Die für die Forschung notwendigen Daten werden regelmäßig als besondere personenbezogene Daten gemäß § 3 Abs. 9 BDSG (Gesundheitsdaten) anzusehen sein, so dass primär die speziellen Forschungsklauseln gemäß §§ 13 Abs. 2 Nr. 6, 28 Abs. 6 Nr. 4 BDSG einschlägig sind. Diese gesetzlichen Forschungsklauseln verzichten jedoch regelmäßig auf eine direkte Ermächtigungsgrundlage für den Umgang mit den besonderen personenbezogenen Daten. Sie normieren stattdessen einen Vorrang der Verwendung anonymisierter und pseudonymisierter Daten und die Einholung einer Einwilligung. Nur wenn der Forschungszweck auf diesen Wegen nicht oder nur mit einem unverhältnismäßigen Aufwand erreicht werden kann, greift die gesetzliche Verwendungserlaubnis.

Wenn klinische Daten für ein Forschungsprojekt erhoben werden, ist im Regelfall die Aufklärung der Patienten und das Einholen einer Einwilligung möglich. Anders sieht es hingegen aus, wenn die Daten bereits zu einem früheren Zeitpunkt im Rahmen einer Behandlung oder eines Forschungsprojekts erhoben wurden und jetzt im Rahmen eines neuen Forschungsvorhabens nachgenutzt werden sollen. Einige Landeskrankenhausgesetze erlauben die Nutzung und Verarbeitung solcher Daten innerhalb der behandelnden Einrichtung auch zum Zwecke der Forschung (z. B. Art. 27 (4) BayKrG). Sollen die Daten aber zum Zwecke der Forschung weitergegeben werden, bieten im Regelfall auch weitgehend formulierte Landeskrankenhausgesetze keine gesetzliche Grundlage. Hier können die speziellen Forschungsklauseln der Datenschutzgesetze greifen, wenn die Übermittlung der Daten zur Durchführung wissenschaftlicher Forschung erforderlich ist, das öffentliche Interesse an dem Forschungsvorhaben das Interesse des Betroffenen an dem Ausschluss der Zweckänderung erheblich überwiegt und der Zweck der Forschung auf andere Weise nicht oder nur mit einem unverhältnismäßigen Aufwand zu erreichen ist. Hierzu ist allerdings konkret zu begründen, warum das Forschungsvorhaben nicht mit anonymen oder pseudonymen Daten umgesetzt werden kann und die Einholung der Einwilligung der betroffenen Patienten unzumutbar ist. In einigen Landesdatenschutzgesetzen ist zusätzlich ein Genehmigungsvorbehalt vorgesehen (z. B. § 33 HDSG). Die Höhe der hierfür gesetzlich vorgeschriebenen Hürden reflektiert zudem die Auflagen der ärztlichen Schweigepflicht nach § 203 Abs. 1 StGB.

Die forschungsspezifischen Datenschutzregelungen lösen somit den regelmäßig bestehenden Konflikt zwischen den konkurrierenden Grundrechten der Forschungsfreiheit gemäß Art. 5 Abs. 3 GG der Daten verarbeitenden Forschungsstellen und dem Recht auf informationelle Selbstbestimmung der Probanden gemäß Art. 2 Abs. 1 und Art. 1 Abs. 1 GG, indem sie über die Einwilligungsmöglichkeit und die strenge Zweckbindung zu einem interessensgerechten Ausgleich der Grundrechte führen.

Im Forschungskontext können weitere untergesetzliche Regelungen gelten, die bestimmte Anforderungen an die Archivierungsfristen für Forschungsdaten stellen. Hier zu nennen wären beispielsweise:

- Richtlinien zur guten wissenschaftlichen Praxis der DFG [5]
- Richtlinien zur guten klinischen Praxis der ICH [6]
- Vorgaben wissenschaftlicher Zeitschriften

Die sich aus diesen Regelungen ergebenden Speicherfristen [vergl. 4] sind jedoch nicht als vorranging gegenüber dem Datenschutzrecht anzusehen. Insofern rechtfertigen sie z. B. nicht den Verzicht auf eine Einwilligung als Rechtsgrundlage für die Verarbeitung. Ggf. sind die Fristen jedoch in den Einwilligungserklärungen aus Transparenzgründen mit zu nennen.

Bezogen auf die oben ausgeführten Nachnutzungsszenarien legen diese Rahmenbedingungen ein dreigleisiges Vorgehen vor:

1. Für bereits bestehende Datensammlungen aus laufenden und abgeschlossenen Forschungsprojekten werden die in den Einwilligungserklärungen notwendigerweise formulierten Zweckbindungen die Nachnutzbarkeit der Daten in der wissenschaftlichen Community regelmäßig stark einschränken oder auch verhindern. Die Einhaltung dieser Restriktionen muss entsprechend von einer technischen Plattform mit unterstützt werden.
2. Parallel ist aber auch zu untersuchen, inwiefern Formulierungen zur einrichtungsübergreifenden Nachnutzung künftig auch in Einwilligungserklärungen neuer Forschungsprojekte mit aufgenommen werden können.
3. Als ergänzende Maßnahme ist die Anonymisierbarkeit einzelner Datenbestände zu prüfen.

3.1 *Grenzen von Einwilligungsszenarien*

Die rechtlich zulässige Verwendung medizinischer Daten, die die informationelle Selbstbestimmung der Patienten wahren muss, kann mit einer Einwilligungserklärung erreicht werden und von wenigen Ausnahmen abgesehen, auch nur so. Diese muss bestimmt sein, so dass klar zu erkennen ist, unter welchen Bedingungen sich die betroffene Person mit der Erhebung, Verarbeitung oder Nutzung welcher Daten einverstanden erklärt. Aus diesem Grund sind weder Blankoeinwilligungen noch pauschal gehaltene Erklärungen, die den Betroffenen die Möglichkeit nehmen, die Tragweite ihres Einverständnisses zu überblicken, ausreichend. Die Anforderungen an die Bestimmtheit sind umso höher, je größer die Tragweite für die Rechte und Freiheiten der betroffenen Person sind. Gemäß § 4a Abs. 3 BDSG bestehen erhöhte Anforderungen an die Bestimmtheit, wenn sich die Verwendung auf besondere Daten im Sinne des § 3 Abs. 9 BDSG bezieht. Somit muss für die Einwilligenden klar erkennbar sein, welche Daten, in welcher Form, von wem, wie lange und wofür verarbeitet oder genutzt werden.

Je konkreter die Einwilligung formuliert ist, desto einschränkender und unter Umständen problematischer ist sie auch für die Forschung, und dies gleich in mehrfacher Hinsicht: Je konkreter der Zweck angegeben wird, desto präziser kann auch der notwendige Datensatz, der für die Verarbeitung erforderliche Personenkreis und die hierfür benötigte Projektlaufzeit bestimmt werden. Im Umkehrschluss geht eine zweckoffenere Erhebung und Speicherung im Regelfall auch mit einer geringeren Einschränkbarkeit des Datenumfangs, einer längeren Vorhaltung der Daten und einem größeren mit deren Verarbeitung betrauten Personenkreis einher.

Dass es in der medizinischen Forschung oft schwer ist, sich auf eine konkret benennbare Fragestellung zu beschränken, ist weithin anerkannt. Häufig wird daher auch akzeptiert, wenn lediglich krankheitsbezogene Einschränkungen gemacht werden. Ausnahmen hierzu

stellen klinische Prüfungen zu Arzneimitteln oder Medizinprodukten dar, die aufgrund der regulatorischen Vorgaben und des geforderten Qualitätsniveaus auf eine Festlegung der Auswertung vor der Datenerhebung angewiesen sind. Aber auch hier kann eine längerfristige Speicherung der wertvollen Daten für zusätzliche Fragestellungen, z. B. zur Generierung neuer Hypothesen, sinnvoll sein. Die Konkretheit der Zweckbezogenheit dient letztlich nur so lange ihrem Ziel der informationellen Selbstbestimmung, wie die Einschränkungen der Forschungsfragestellung von der Mehrheit der Patienten auch nachvollzogen und verstanden werden können. Somit kann auch das Gebot der Verständlichkeit der Einwilligungserklärung schon eine Aufweichung des Prinzips der möglichst engen Definition der Zweckbezogenheit bedeuten.

Während eine in Grenzen zweckoffene Erhebung, Speicherung und Verarbeitung medizinischer Daten dem Prinzip einer informierten Einwilligung häufig nicht direkt entgegensteht, sind die damit regelmäßig verbundenen Verschiebungen der organisatorischen Rahmenbedingungen gesondert zu betrachten. Die klare Unterscheidbarkeit unterschiedlicher organisatorischer Vorgaben in den Augen der Patienten, wie z. B. die Zeitdauer der Speicherung (s. o.), empfiehlt diese für die Berücksichtigung in einer abgestuften Einwilligungserklärung [7]. Auch die Methodik der abgestuften Einwilligungserklärung führt jedoch nicht automatisch zu einer ausreichenden Wahrnehmung, bzw. einem informierten Verständnis aller Risiken bei den Patienten.

Mit der verstärkt wahrgenommenen Bedeutung der Biobank-gestützten Forschung in den letzten Jahren ist die Diskussion um die nötige Enge der Zweckbezogenheit der Einwilligung unter dem Stichwort „broad consent" erneut und kontrovers geführt worden. Der Deutsche Ethikrat hat in einer Empfehlung zu Humanbiobanken für die Forschung gar gesetzlich geregelte Rahmenbedingungen, wie z. B. ein Biobankgeheimnis, gefordert, welche eine zweckoffene Einwilligung ermöglichen sollen [8]. Da bisher jedoch völlig offen ist, ob und in welcher Form der Gesetzgeber diesen Vorschlag aufnimmt, bleibt in jedem Einzelfall zu prüfen und abzuwägen, ob der Grad der Bestimmtheit einer Einwilligung den Forschungsinteressen und der Informiertheit der Probanden noch gerecht wird. Für Letzteres spielt auch die Wahrnehmbarkeit der durch die längerfristige, vergleichsweise zweckoffene und breit nutzbare Speicherung medizinischer Daten entstehenden Risiken eine entscheidende Rolle. Um diese auf ein vertretbares Maß zu reduzieren, sind entsprechende technische und organisatorische Maßnahmen, flankiert durch klar geregelte rechtliche Verantwortlichkeiten, zu implementieren. Als Blaupause hierfür können die generischen Datenschutzkonzepte der TMF herangezogen werden, die auf nationaler Ebene mit den zuständigen Arbeitskreisen der Konferenz der Datenschutzbeauftragten des Bundes und der Länder abgestimmt wurden [9]. Diese sehen einen zum Zeitpunkt der Datenerhebung feststehenden starren rechtlichen und organisatorischen Rahmen für die Verarbeitung und Nutzung der Daten vor, der im Sinne einer informierten Einwilligung auch in der Patienteninformation- und Einwilligungserklärung dargestellt ist. Innerhalb dieses organisatorischen und rechtlichen Rahmens ist jedoch zu einem späteren Zeitpunkt die Bereitstellung von Daten zu vorher noch nicht festlegbaren Fragestellungen an vorher noch unbekannte Forschergruppen möglich. Hierfür muss jedoch das ursprünglich festgelegte Verfahren eingehalten werden, welches z. B. die Prüfung eines Da-

tennutzungsantrags durch ein fest definiertes Gremium (z. B. „Ausschuss Datenschutz") vorsieht.

3.2 *Anonymisierung*

Ebenfalls zu prüfen ist, inwiefern Daten für die Nachnutzung anonymisiert werden können. Daten sind nach § 3 Abs. 6 BDSG anonymisiert, wenn sie entweder „nicht mehr" oder „nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können". Während die erste Option als absolute Anonymisierung bezeichnet wird, ist die letztere realistischerweise die häufigere und wird mit dem Begriff der faktischen Anonymisierung belegt [10]. Grundsätzlich wird von einer Anonymisierung ausgegangen, wenn identifizierende und medizinische Daten getrennt werden, keine Zuordnungsregel mehr existiert und anhand der medizinischen Daten allein keine Reidentifizierung möglich ist. Anonymisierte Daten gelten damit für nicht mehr personenbeziehbar, so dass für sie auch das Datenschutzrecht samt Einwilligungsvorbehalt nicht mehr anzuwenden ist [11, S. C35]. Problematisch im Umgang mit anonymisierten Daten ist, dass sich der Status der Anonymität im Laufe der Zeit ändern kann, z. B. wenn ein Nutzer der Daten aufgrund einer bestimmten Kombination medizinischer und sozialer Daten auf die Identität des zugehörigen Patienten schließen kann. In diesem Falle würde es sich wieder um personenbezogene Daten handeln, die entsprechend der Vorschriften der Datenschutzgesetze des Bundes und der Länder zu behandeln wären. Problematisch an einem solchen Szenario ist, dass sich im Vorfeld nicht immer ausreichend präzise einschätzen lässt, ob und wann ein solcher Fall eintreten kann. Zur Vorbeugung wird daher empfohlen, auch anonymisierte Datenexporte nur zweckbezogen an definierte Nutzerkreise abzugeben und insbesondere auf die freie Verfügbarmachung medizinischer Daten im Internet in so genannten Public-Use-Files zu verzichten. Eine größere Sicherheit zur Verhinderung solcher Reidentifizierungen kann durch eine k-Anonymisierung erreicht werden [12], wobei jedoch zu klären ist, für welche Anwendungsfälle und Nachnutzungsszenarien solcher Art modifizierte Datenbestände noch adäquat nutzbar sind.

Bezogen auf die hier relevanten Anwendungsfälle kann eine Anonymisierbarkeit genetischer Daten grundsätzlich in Zweifel gezogen werden. Möglich wäre diese aber ggf. für bestimmte Bilddaten, wie es z. B. auch für Referenzdatenbanken z.T. schon heute durchgeführt wird[6].

4 Verantwortlichkeit

Die Speicherung und Verarbeitung sensibler medizinischer Daten setzt eine juristisch belastbare und für jeden Betroffenen nachvollziehbare Regelung der Verantwortlichkeit voraus. Bei der Planung einer langfristigen und einrichtungsübergreifenden Infrastruktur für die Nachnutzung von Forschungsdaten ist von vornherein auch zu überlegen, welche verlässliche und vertrauenswürdige Institution mit ebenfalls langfristiger Perspektive die Verantwortung im juristischen Sinne übernehmen kann. Dies kann ein von den beteiligten Institutionen gegründeter Verein sein, es sind aber auch andere Lösungen und Formen, die als juristische Person ansprechbar sind, denkbar. Bei der Initiierung einer Infrastruktur aus einem geförder-

[6] siehe z.B. das Medical Imaging Resource Center (MIRC) der RSNA unter http://mirc.rsna.org

ten Forschungsprojekt heraus ist auch an Regelungen nach Auslaufen der Förderung zu denken. In jedem Fall sollte auch darüber nachgedacht werden, ob die Regelung einer möglichen Rechtsnachfolge für die zunächst verantwortliche Institution sinnvoll ist. Die notwendige Transparenz gegenüber den beteiligten Institutionen wie den Betroffenen im datenschutzrechtlichen Sinne erfordert die verständliche Darlegung der Verantwortlichkeiten. Dies betrifft in Forschungsprojekten regelmäßig auch schon die Einwilligungserklärungen.

Der verantwortlichen Institution wird empfohlen, ein Gremium zu schaffen, welches für datenschutzrechtliche Fragen und Entscheidungen zuständig ist. Bei der Besetzung dieses Gremiums (z. B. als „Ausschusses Datenschutz" benannt) ist darauf zu achten, dass Interessenskonflikte soweit wie möglich vermieden werden. Das Gremium sollte neben der Beratung einzelner Entscheidungen, z. B. welcher Anfrage nach Daten stattgegeben wird, auch für die Ausarbeitung und Fortschreibung der datenschutzrechtlich relevanten Regelwerke und Policies verantwortlich sein. Das Aufgabengebiet des Ausschuss Datenschutz kann z. T. überlappend mit jenem eines betrieblichen oder behördlichen Datenschutzbeauftragten der beteiligten Einrichtungen oder auch eines übergreifenden Vereins sein. Wenn die übergreifende Institution über einen eigenen Datenschutzbeauftragten verfügt, sollte dieser entsprechend auch Mitglied des Ausschuss Datenschutz werden. In Bezug auf die langfristige pseudonymisierte Aufbewahrung von Gesundheitsdaten kommt dem Ausschuss Datenschutz jedoch eine besondere Verantwortlichkeit zu, die im Regelfall eine Besetzung mit mehreren Personen mit ausreichender Sachkenntnis empfehlenswert erscheinen lässt. Somit sollten auch alle relevanten Entscheidungen mindestens nach dem Vier-Augen-Prinzip getroffen werden.

5 Daten und Datenkategorien

Aufgrund des generischen Ansatzes des hier entwickelten Konzepts, können keine konkreten Daten und Datenkategorien benannt werden. In einem hiervon abzuleitenden konkreten Datenschutzkonzept sind diese jedoch aufzuführen. Allgemein sind dabei, bezogen auf die im Rahmen des Projekts LABIMI/F untersuchten Use Cases (s. Kap. 2), die verschiedenen Bildarten unterschiedlicher Modalitäten samt Metadaten und ggf. begleitenden klinischen Daten zu berücksichtigen. Für den Use Case der Genomdaten ist das gesamte genomische Datenspektrum von der Keimbahn-DNA bis zu metabolischen Daten zu berücksichtigen. Zudem sind ggf. Daten unterschiedlicher technischer Vorverarbeitungen und Auswertungsmethoden in die Betrachtung einzubeziehen. Nicht zu vergessen ist auch für diesen Use Case die Notwendigkeit ausführlicher klinischer Daten zur Beschreibung des zu den genetischen Daten zugehörigen Phänotyps. Als letzte wichtige Datenkategorie sind für beide Use Cases die Identitätsdaten der betroffenen Patienten und Probanden zu nennen, die ggf. auch verschiedene pseudonyme Zuordnungen umfassen können.

6 Technische und organisatorische Schutzmaßnahmen

Die Nutzung einer Infrastruktur zur Nachnutzung von Forschungsdaten kann die Ziele von Datenschutzmaßnahmen in verschiedener Hinsicht gefährden. Naheliegend ist die Gefährdung der Vertraulichkeit, wenn personenbezogene Daten einrichtungsübergreifend verarbeitet und genutzt werden. Weniger offensichtlich aber nicht minder relevant sind die Schutzziele Verfügbarkeit, Integrität, Transparenz und Revisionsfähigkeit der Daten bzw.

Datenverarbeitung. Eine beispielhafte Beschreibung dieser Schutzziele findet sich u. a. auch in der Orientierungshilfe Cloud Computing der Konferenz der Datenschutzbeauftragten des Bundes und der Länder [13, S. 8].

Die technischen und organisatorischen Maßnahmen dienen der Umsetzung und Gewährleistung dieser Schutzziele. Welche konkreten Maßnahmen getroffen werden müssen, wird im Regelfall von den zuständigen Datenschutzbeauftragten nicht vorgeschrieben, zumal diese auch dem steten technischen Wandel unterliegen. Ausnahmen stellen die modellhaften Lösungen zum Datenschutz in der medizinischen Forschung der TMF dar, die mit den Datenschutzbeauftragten des Bundes und der Länder abgestimmt wurden [9], die jedoch noch keine Vorgaben für den Aufbau einer einrichtungsübergreifenden Infrastruktur zur Nachnutzung von Forschungsdaten enthalten. Wichtig ist in jedem Fall die Einhaltung des Stands der Technik hinsichtlich der getroffenen Schutzmaßnahmen. Hilfestellungen zur Einhaltung eines Grundschutzes nach aktuellem Stand der Technik bietet das Bundesamt für Sicherheit in der Informationstechnik (BSI)[7]. Um sich an dem Stand der Technik zu orientieren, wären ggf. auch die Empfehlungen für Cloud-Computing von Interesse. Ebenfalls vom BSI gibt es ein umfassendes Eckpunktepapier zu Sicherheitsempfehlungen für Cloud-Anbieter [14].

Eine wichtige Maßnahme zur Sicherung der Vertraulichkeit von Daten ist die Pseudonymisierung. Eine effektive Pseudonymisierung, verbunden mit einer informationellen Gewaltenteilung gemäß den Grundprinzipien der generischen TMF-Datenschutzkonzepte [9], sollte im Forschungskontext als Mindeststandard umsetzbar sein, zumindest was den für eine Nachnutzung relevanten Datenbestandteil anbetrifft. Dies gebietet schon das datenschutzrechtliche Grundprinzip der Datensparsamkeit (§ 3a BDSG).

Grundsätzlich gilt, dass alle Verarbeitungsschritte und Zugriffe auf die Daten nach dem Stand der Technik zu protokollieren sind. Die Protokolldateien selbst sind geschützt aufzubewahren. Entsprechend müssen die eingesetzten Systeme auch über einen effektiven Zugangsschutz verfügen, der idealerweise nicht nur in einer Autorisierung mit Hilfe eines Benutzernamens und eines Passwortes besteht. Zudem sind im Sinne der datenschutzrechtlich vorgegebenen Datensparsamkeit granulare Zugriffsberechtigungen zu vergeben. Über offene Netze (Internet) übermittelte sensible Gesundheitsdaten sind in jedem Fall nach dem Stand der Technik zu verschlüsseln und somit vor einer unautorisierten Einsicht zu schützen. Zusätzlich sind die Daten nur pseudonymisiert zu übermitteln, so dass identifizierende und medizinische Daten nicht gemeinsam übertragen werden.

Aus den bis hierher aufgeführten Rahmenbedingungen ergeben sich für eine technische und organisatorische Umsetzung einer LZA-Infrastruktur drei Modellvarianten:

1. Anonyme Bereitstellung von biomedizinischen Forschungsdaten zur wissenschaftlichen Nachnutzung
2. Zentrale Bereitstellung von Metadaten als Vermittlungsportal zur Nachnutzung der in dezentralen Archiven gelagerten biomedizinischen Forschungsdaten

[7] siehe https://www.bsi.bund.de/gshb

3. Zentrale Archivierung und Bereitstellung biomedizinischer Forschungsdaten auf Basis einer informierten Einwilligung

Die drei nachfolgend vorgestellten Modellvarianten können auch kombiniert zur Anwendung kommen. Dies kann insbesondere für heterogene Datenbestände und bei Beteiligung von Forschungseinrichtungen mit unterschiedlichen technischen Voraussetzungen und wissenschaftlichen Zielsetzungen von Interesse sein.

6.1 *Modell 1: Anonyme Bereitstellung von Forschungsdaten*

Wie schon weiter oben ausgeführt, wird die anonyme Bereitstellung der Bild- und Genomdaten die einzige Variante der zentralen Bereitstellung von Forschungsdaten zur Nachnutzung sein, wenn keine Einwilligung als Rechtsgrundlage eingeholt werden kann. Dies wird regelmäßig für Altdaten gelten. Allerdings wird eine anonyme Aufbewahrung der Forschungsdaten nicht den Anforderungen der Nachvollziehbarkeit genügen, die wesentlich die gesetzlichen wie untergesetzlichen Archivierungspflichten begründen. Somit ließen sich mit einer anonymisierten Datensammlung lediglich die Use Cases zur wissenschaftlichen Nachnutzung umsetzen. Die synergistische Nutzung einer einzigen Infrastruktur sowohl für die verpflichtende Archivierung als auch die wissenschaftliche Nachnutzung der Daten ist mit einem solchen Anonymisierungsmodell nicht umsetzbar.

Weiterhin ist zu beachten, dass vor allem genetische Daten, oft aber auch Bilddaten, kaum noch oder nur in Ausnahmefällen effektiv anonymisierbar sind. Somit ist ein solches Lösungsmodell nur für ausgewählte Datenarten möglich, die dem Kriterium der Anonymisierbarkeit genügen. Dies schränkt den potentiellen Nutzen einer solchen Infrastruktur weiter ein.

Wichtig ist weiterhin, dass die Anonymisierung der Daten selbst eine Verarbeitung personenbezogener Daten im Sinne des Datenschutzrechts darstellt, die einer Rechtsgrundlage bedarf. Dies bedeutet, dass die Anonymisierung auf jeden Fall dezentral in den jeweiligen Einrichtungen zu erfolgen hat, die die Daten rechtmäßig erhoben haben. Weiterhin ist darauf zu achten, dass die Daten im Rahmen der Anonymisierung keinem erweiterten Personenkreis offenbart werden. Demnach sind interaktive Prozesse, die z. B. im Rahmen eines Anonymisierungsworkflows für Bilder nötig sein können, nur von jenem Nutzerkreis auszuführen, der rechtmäßigen Zugriff auf die Daten hat. Vor dem Hintergrund der heute in Krankenhäusern restriktiv geregelten Zugriffsrechte [15], insbesondere für administratives Personal, ist dies nicht immer einfach umsetzbar.

6.2 *Modell 2: Vermittlungsportal auf der Basis von Metadaten*

Wie auch das erste Modell adressiert auch die hier dargestellte Variante eines Vermittlungsportals auf der Basis von Metadaten lediglich Anwendungsfälle der wissenschaftlichen Nachnutzung der Daten. Für die Archivierung der Daten zum späteren Nachweis korrekten Arbeitens kommt das Modell ebenfalls nicht in Betracht.

In den beteiligten Einrichtungen werden standardisierte Metadaten zu den Bildern und Genomdaten erfasst und in pseudonymer Form vorgehalten. Diese pseudonymen Metadaten

werden in anonymer Form exportiert und dann einer zentralen Stelle übermittelt. Dort stehen sie für Recherchen durch Wissenschaftler frei zur Verfügung. Die Ergebnisse der Recherchen verweisen dann auf interessierende Primärdaten an bestimmten beteiligten Stellen, mit denen der recherchierende Forscher dann jeweils separat eine mögliche weitere Nutzung der Daten klären kann. Die interessierenden Datensätze können auf der Basis der lokal vorgehaltenen pseudonymen Metadaten und der zentral genutzten Abfrage jeweils durch Beschäftigte der beteiligten dezentralen Stellen ermittelt und ggf. für die weitere Forschung zur Verfügung gestellt werden. Dieses Prinzip der Vermittlung anhand zentral vorgehaltener anonymer Daten orientiert sich an dem im CRIP-Projekt (vormals RZPD) entwickelten Verfahren [16].[8]

Ein wesentlicher Vorzug dieses Verfahrens ist es, dass an zentraler Stelle ebenfalls keine personenbezogenen Daten vorgehalten werden müssen, so dass für die Übermittlung der Daten von den beteiligten Stellen an die zentrale Plattform keine Einwilligungen als Rechtsgrundlage notwendig sind. Lediglich die Einrichtung, aus der die Metadaten stammen, muss in der zentralen Datenbank gespeichert sein. Dazu kommt, dass sich die Metadaten zu Bildern und Genomdaten deutlich besser anonymisieren lassen als diese selbst. Zu diesen Metadaten könnten auch klinische bzw. phänotypische Daten gehören, die für viele Anwendungsfälle (s. Kapitel 2) hoch relevant sind. Um eine effektive Anonymisierung zu erreichen, können die zentral für die Recherche bereitgestellten Metadaten auch k-anonymisiert[9] werden. Allerdings ist dann zu klären, inwieweit die zentral recherchierten k-anonymisierten Metadaten mit den Rechercheergebnissen aus den lokalen pseudonymisierten Metadatenbeständen übereinstimmen.

Eine Gefahr dieses Modells besteht jedoch darin, dass unnötige Anfragen produziert werden können, wenn die zu den zentral recherchierten Metadaten zugehörigen Primärdaten für eine Herausgabe an externe Forscher gar nicht bereitstehen. Daher wird empfohlen, die zu den Primärdaten gehörenden Policies, insofern sie die wissenschaftliche Nachnutzung betreffen, auch schon in den Metadaten mit zu kodieren. Anders herum kann aber die bei den beteiligten Institutionen verbleibende Entscheidungshoheit über die Nutzung der Primärdaten auch als Vorteil gesehen werden.

Eine zentrale Infrastruktur kann in diesem Modell auch die Anfragen der in der zentralen Datenbank recherchierenden Forscher an die beteiligten Einrichtungen übermitteln. Dabei können die beteiligten Einrichtungen, die über passende Bild- oder Genomdaten verfügen, dem anfragenden Forscher auch verheimlicht werden, so dass sie sich erst bei Interesse an einer Kooperation offenbaren müssen. In diesem Falle müsste von dem anfragenden Forscher eine Beschreibung seines Forschungsprojekts bereitgestellt und mit übermittelt werden.

Weiterhin wäre auch eine technische Unterstützung der Übermittlung der eigentlichen Daten an den externen Forscher durch die zentrale Infrastruktur möglich. Diese könnte den Transfer z. B. durch die Bereitstellung einer zentralen PKI absichern und vereinfachen. Dabei dürfen die Daten jedoch der zentralen Stelle nicht unnötig offenbart werden. Eine weitere große

[8] Ebenfalls auf dem Prinzip des CRIP-Portals basiert das Projektportal im Deutschen Biobanken-Register (www.biobanken.de)
[9] Die TMF stellt ein Softwareframework für die k-Anonmyisierung kostenfrei bereit, siehe http://www.tmf-ev.de/Produkte/P100201

Hilfestellung für die nachnutzenden Forscher und die beteiligten datenliefernden Einrichtungen wäre die Bereitstellung passender Musterverträge und Nutzungsvereinbarungen.

Auch wenn sich bestimmte wissenschaftliche Fragestellungen möglicherweise sogar allein anhand der anonymen Metadaten beantworten lassen, so hängt doch die eigentliche Nachnutzung der Daten samt ihres potentiellen wissenschaftlichen Erkenntnisgewinns davon ab, dass die lokal gespeicherten Primärdaten auch tatsächlich für externe Forscher zu einem späteren Zeitpunkt bereitgestellt werden können. Somit ist auch für dieses Modell die Entwicklung und Nutzung einer ausreichend offen formulierten Einwilligungserklärung eine zentrale Voraussetzung. Allerdings kann diese Modellvariante den beteiligten Einrichtungen und deren Probanden eine maximale Flexibilität hinsichtlich der weiteren Nutzungsbestimmungen gewähren und gleichzeitig auch eine weitgehende Verwertungshoheit garantieren.

Mit der Sicherheit der beteiligten datenliefernden Stellen, dass die wertvollen Daten in ihrem eigenen Gewahrsam verbleiben, bis zu einer konkreten Nachnutzung mit einem konkreten Forscher eine vertragliche Vereinbarung geschlossen wird, ist jedoch auch der erhöhte technische Aufwand der sicheren langfristigen Archivierung und Vorhaltung der Daten verbunden.

6.3 *Modell 3: Zentrale Archivierung und Bereitstellung von Forschungsdaten*

Das Ziel der dritten Modellvariante ist es, durch die Bereitstellung einer zentralen Infrastruktur die beteiligten Einrichtungen von den technischen und organisatorischen Aufwänden für die sichere langfristige Archivierung der Forschungsdaten zu entlasten. Gleichzeitig soll eine einzige Infrastruktur sowohl die Anwendungsfälle der wissenschaftlichen Nachnutzung unterstützen, wie auch der Dokumentationspflicht zum Zwecke der Nachvollziehbarkeit der Forschungsarbeiten genügen.

Das Modell sieht vor, dass nach Abschluss der Forschungsprojekte in den beteiligten Einrichtungen die Forschungsprimärdaten, mit standardisierten Metadaten versehen, vollständig in ein zentrales Archiv transferiert werden. Gemäß den generischen Datenschutzlösungen der TMF wäre für ein solches zentrales Repository eine neue Pseudonymisierung vorzusehen [vergl. Modell B in 9]. Der Zusammenhang zwischen den lokalen Pseudonymen und dem zentralen Pseudonym sollte idealerweise alleinig bei einer vertrauenswürdigen dritten Stelle (TTP, Trusted Third Party) hinterlegt sein. Da die Daten jedoch von den dezentralen Einrichtungen in pseudonymer und damit nach mehrheitlicher Rechtsauffassung der zuständigen Aufsichtsbehörden in Deutschland immer noch in personenbezogener bzw. personenbeziehbarer Form übermittelt werden, ist eine Rechtsgrundlage hierfür erforderlich[10]. Dieses Modell sieht zudem eine vollständige Übermittlung der genomischen und Bilddaten vor, auch im Sinne einer herkömmlichen Archivierung, so dass für die Bewertung der Personenbeziehbarkeit auch das hohe Reidentifikationspotential der Daten selbst mit zu berücksichtigen ist. Insofern basiert dieses Modell zwingend auf einer entsprechend offen formulierten Einwilligung (s. Kap. 3.1).

[10] eine moderatere Auffassung vertritt Scholz in [17, Rdnr. 218-219]

Der Zugriff der Forscher auf diese zentral gespeicherten Daten hat strikten Regeln zu folgen, die bereits zum Zeitpunkt der Datenerhebung feststehen und in der Patienteninformation und Einwilligungserklärung dokumentiert sein sollten. Dies umfasst z. B. die Prüfung eines Antrags auf Recherche und Nutzung der Daten durch ein unabhängiges und fachlich kompetent besetztes Gremium. Im Kontext der generischen Datenschutzkonzepte der TMF hat sich für ein solches Gremium die Bezeichnung „Ausschuss Datenschutz" etabliert. Nach erfolgreicher Prüfung des Antrags eines Forschers, bekommt dieser die Möglichkeit einer Recherche in den Metadaten.

Für den Export vollständiger Datensätze für die weitere Auswertung im Rahmen eines genehmigten Projekts ist zu spezifizieren, welche Daten benötigt werden und ob im Rahmen der Auswertung möglicherweise mit relevanten Ergebnissen für einzelne Patienten zu rechnen ist. Wenn keine relevante individuelle Rückmeldung der Ergebnisse an die Patienten zu erwarten ist, werden die Daten für den Export anonymisiert, andernfalls werden die Daten mit einem neuen Pseudonym versehen und exportiert. Sowohl bei anonymisierten wie auch pseudonymisierten Exporten ist darauf zu achten, dass sich die Sortierreihenfolge der exportierten Datensätze nicht nach dem langfristigen Pseudonym im zentralen Archiv richtet, sondern z. B. nach den neu erzeugten anonymen oder pseudonymen IDs.

Im Gegensatz zu dem vorhergehend beschriebenen Modell 2 setzt die hier beschriebene Architekturvariante darauf, dass es eine grundlegende Bereitschaft zum Datenaustausch aller beteiligten Einrichtungen gibt und dass bestimmte Grundregeln für diesen Datenaustausch auch im Vorhinein festgelegt werden können. Damit entfällt die Notwendigkeit, für jeden Datenexport bei den ursprünglichen datenliefernden Stellen nachfragen zu müssen. Die Verantwortung für die weitere Nachnutzung der Daten wird damit an eine übergeordnete Einrichtung delegiert, auch wenn für Einzelfälle natürlich ein Rückfragemanagement implementiert werden kann. Hierfür ist auch wiederum sicherzustellen, dass ggf. vorhandene Nutzungseinschränkungen oder zusätzliche Policies im zentralen Archiv bekannt sind und für die Recherche in den Metadaten schon berücksichtigt werden können.

Eine offene und im Vorfeld einer Implementierung noch zu klärende Frage betrifft die aus der Perspektive des Datenschutzes gewünschte doppelte Pseudonymisierung der im zentralen Archiv langfristig vorzuhaltenden Daten. Jenseits der datenschutzrechtlichen Begründung dieser Maßnahme ist zu fragen, ob diese Veränderung der Daten mit der eigentlichen Archivierungspflicht, die ja typischerweise eine unveränderliche Speicherung erfordert, in Einklang zu bringen ist. Hier wird ein möglicher Zielkonflikt deutlich, der beim Aufbau einer LZA-Infrastruktur vorliegen kann. Für eine wissenschaftliche Nachnutzung der medizinischen Daten ist aus datenschutzrechtlicher Sicht eine Minimierung des Reidentifizierungspotentials und ggf. auch eine entsprechende Veränderung oder Vergröberung der Daten gewünscht und häufig auch möglich. Für die zweite Zielsetzung einer solchen LZA-Infrastruktur hingegen, die gerade darin besteht, zu einem späteren Zeitpunkt ein Forschungsprojekt möglichst genau nachvollziehen zu können, ist es essentiell, die Originaldaten möglichst unverändert aufzubewahren.

7 Fristen

Die Fristen für die Aufbewahrung personenbezogener Daten richten sich nach den entsprechenden rechtlichen Grundlagen. Für die eigentliche Archivierung der wissenschaftlichen Daten zum Zwecke der nachvollziehbaren Dokumentation eines Forschungsprojekts gelten die entsprechend bereits in Kap. 3 aufgeführten gesetzlichen und untergesetzlichen Bestimmungen. Die wissenschaftliche Nachnutzung der Daten ist hingegen nicht eigenständig gesetzlich geregelt und unterliegt damit den allgemeinen Bestimmungen des Datenschutzes. Der dort verankerte Grundsatz der Datensparsamkeit gebietet, dass die Daten nur so lange in personenbezogener Form aufbewahrt werden, wie hierfür auch eine Notwendigkeit besteht. Darüber hinaus erfordert die informierte Einwilligung als Rechtsgrundlage, dass die Betroffenen über die Fristen aufgeklärt werden. Somit ist gerade für die wissenschaftliche Nachnutzung in Abhängigkeit von dem Datenmaterial und den zu unterstützenden Use Cases genau festzulegen, über welche Zeiträume eine pseudonyme Vorhaltung der Daten, entweder lokal oder zentral, erforderlich ist.

8 Anhang

8.1 *Abkürzungsverzeichnis*

AK	Arbeitskreis
AMG	Gesetz über den Verkehr mit Arzneimitteln – Arzneimittelgesetz
AWMF	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (www.awmf.org)
BayKrG	Bayerisches Krankenhausgesetz
BDSG	Bundesdatenschutzgesetz
BSI	Bundesamt für Sicherheit in der Informationstechnik (www.bsi.de)
CRIP	Central Research Infrastructure for molecular Pathology (www.crip.fraunhofer.de)
CT	Computer-Tomografie
dbGaP	NIH Database of Genotypes and Phenotypes (www.ncbi.nlm.nih.gov/gap)
DFG	Deutsche Forschungsgemeinschaft (www.dfg.de)
DGEpi	Deutsche Gesellschaft für Epidemiologie e. V. (http://dgepi.de)
DNA	Deoxyribonucleic acid (Desoxyribonukleinsäure)
DSK	Datenschutzkonferenz – Konferenz der Datenschutzbeauftragten des Bundes und der Länder
EBI	European Bioinformatics Institute, Teil des EMBL (www.ebi.ac.uk)
EMBL	European Molecular Biology Laboratory (www.embl.org)
ENA	European Nucleotide Archive am EMBL-EBI (www.ebi.ac.uk/ena)
fMRI	functional Magnetic Resonance Imaging (funktionelle Kernspintomographie)
GenBank	Datenbank genetischer Sequenzdaten des NCBI (www.ncbi.nlm.nih.gov/genbank)
GG	Grundgesetz der Bundesrepublik Deutschland
GWAS	Genomweite Assoziationsstudie(n)
HDSG	Hessisches Datenschutzgesetz
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (www.ich.org)
ID	Identifikationsnummer
k-Anonymisierung	Verfahren zur Anonymisierung einer Datensammlung, so dass jede Merkmalskombination, die potentiell für einen reidentifizierenden Abgleich genutzt werden könnte, in mindestens k Datensätzen vorkommt
KIS	Krankenhausinformationssystem
LABIMI/F	Langzeitarchivierung biomedizinischer Forschungsdaten; von der DFG gefördertes Projekt (www.labimi-f.med.uni-goettingen.de)
LDSG	Landesdatenschutzgesetz
LZA	Langzeitarchivierung
MBO	Musterberufsordnung für Ärzte
MIRC	Medical Imaging Resource Center, von der RSNA initiiertes Open Source Software-Projekt (https://rsna.org/MIRC.aspx)
MPG	Gesetz über Medizinprodukte - Medizinproduktegesetz
MRT	Magnetresonanztomographie
NCBI	US National Center for Biotechnology Information (www.ncbi.nih.gov)
NIH	US National Institutes of Health (www.nih.gov)
PDB	Protein Data Bank der RCSB (www.rcsb.org/pdb)
PET	Positronenemissionstomographie

PKI	Public Key Infrastruktur
RCSB	Research Collaboratory for Structural Bioinformatics; Zusammenschluss von Forschungseinrichtungen im Bereich der Bioinformatik (http://home.rcsb.org)
RöV	Verordnung über den Schutz vor Schäden durch Röntgenstrahlen - Röntgenverordnung
RSNA	Radiological Society of North America (www.rsna.org)
RZPD	Deutsches Ressourcenzentrum für Genomforschung GmbH (www.rzpd.de)
SGB	Sozialgesetzbuch
SLA	Service Level Agreement
StGB	Strafgesetzbuch
StrlSchV	Verordnung über den Schutz vor Schäden durch ionisierende Strahlen – Strahlenschutzverordnung
TMF	TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (www.tmf-ev.de)
TTP	Trusted Third Party
UAG	Unterarbeitsgruppe

8.2 *Literatur*

1. Mailman, M.D., Feolo, M., Jin, Y. et al., *The NCBI dbGaP database of genotypes and phenotypes.* Nat Genet, 2007. **39**(10): S. 1181-1186.

2. Krawczak, M., Goebel, J.W., Cooper, D.N., *Is the NIH policy for sharing GWAS data running the risk of being counterproductive?* Investig Genet, 2010. **1**(1): S. 3.

3. Dierks, C. *Rechtsgutachten zur elektronischen Archivierung. Teil 2: Spezifische Rechtsfragen zur elektronischen Aufbewahrung von Dokumenten und Dateien in klinischen Studien*. 2010. TMF, http://www.tmf-ev.de/produkte/P042011 (Abruf: 2012-08-08).

4. Mathieu, N., Lönhardt, B., Grütz, R., Weil, P., Drepper, J., Krawczak, M., *Ethische und rechtliche Implikationen der Speicherung humaner Genomdaten.* medizinische genetik, 2013. **25**(2): S. 278-283.

5. DFG, *Vorschläge zur Sicherung Guter Wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft"; Denkschrift*. 1998, Weinheim: Wiley-VCH.

6. ICH *Guideline for Good Clinical Practice*. 1996. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use E6(R1), http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1__Guideline.pdf.

7. Harnischmacher, U., Ihle, P., Berger, B., Goebel, J.W., Scheller, J., *Checkliste und Leitfaden zur Patienteneinwilligung*. 2006, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

8. Ethikrat *Humanbiobanken für die Forschung*. 2010. Deutscher Ethikrat, http://www.ethikrat.org/dateien/pdf/stellungnahme-humanbiobanken-fuer-die-forschung.pdf (Abruf: 2013-04-19).

9. Reng, C.-M., Pommerening, K., Specker, C., Debold, P., *Generische Lösungen zum Datenschutz für die Forschungsnetze in der Medizin: Im Auftrag des Koordinierungsrates der Telematikplattform für Medizinische Forschungsnetze*. 1., Auflage Aufl. 2006, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

10. Dammann, U., *§3 Weitere Begriffsbestimmungen*, in *Bundesdatenschutzgesetz*, Hrsg.: S. Simitis. 2006, Nomos: Baden-Baden. S. 263.

11. Roßnagel, A., Hornung, G., Jandt, S. *Rechtsgutachten zum Datenschutz in der medizinischen Forschung. Teil 2: Gutachten zur Mitnutzung von Versorgungsdaten und zur elektronischen Gesundheitskarte nach §291a SGB V*. 2009. TMF, http://www.tmf-ev.de/produkte/P039031.

12. Sweeney, L., *k-Anonymity: A model for protecting privacy.* International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002. **10**(05): S. 557-570.

13. DSK *Orientierungshilfe - Cloud Computing*. 2011. AK Technik und Medien: Konferenz der Datenschutzbeauftragten des Bundes und der Länder, http://www.datenschutz-bayern.de/technik/orient/oh_cloud.pdf (Abruf: 2012-08-08).

14. BSI *Eckpunktepapier Sicherheitsempfehlungen für Cloud Computing Anbieter - Mindestanforderungen in der Informationssicherheit*. 2012. Bundesamt für Sicherheit in der Informationstechnik, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Mindestanforderungen/Eckpunktepapier-Sicherheitsempfehlungen-CloudComputing-Anbieter.pdf.

15. DSK *Orientierungshilfe Krankenhausinformationssysteme*. 2011. UAG Krankenhausinformationssysteme der AK Gesundheit und Soziales, Technische und organisatorische Datenschutzfragen: Konferenz der Datenschutzbeauftragten des Bundes und der Länder, http://www.datenschutz-kirche.de/KIS (Abruf: 2012-08-08).

16. Gros, O., Heidtke, K.R., Schröder, C., *CRIP Privacy Regime und IT-Architektur - ein Modell für Meta-Biobanken*, in *56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), 6. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi)*, 2011: Mainz, http://www.egms.de/static/en/meetings/gmds2011/11gmds524.shtml.

17. Scholz, P., *§3 Weitere Begriffsbestimmungen. Rdnr. 212-222, 266-278*, in *Bundesdatenschutzgesetz*, Hrsg.: S. Simitis. 2011, Nomos: Baden-Baden.