



# IT-Infrastrukturen in der patientenorientierten Forschung

Aktueller Stand und Handlungsbedarf **2016**

verfasst und vorgelegt vom

IT-Reviewing-Board der TMF – Technologie- und Methodenplattform

für die vernetzte medizinische Forschung e.V.



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





# IT-Infrastrukturen in der patientenorientierten Forschung

Aktueller Stand und Handlungsbedarf

2016

verfasst und vorgelegt vom  
IT-Reviewing-Board der TMF –Technologie- und Methodenplattform  
für die vernetzte medizinische Forschung e.V.



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



# Impressum

## **IT-Infrastrukturen in der patientenorientierten Forschung**

### **Aktueller Stand und Handlungsbedarf – 2016**

#### **Redaktion & Herausgabe**

Dr. Johannes Drepper, Sebastian C. Semler

TMF – Technologie- und Methodenplattform  
für die vernetzte medizinische Forschung e. V.  
Charlottenstraße 42 | D-10117 Berlin  
Tel.: +49 (0)30 22 00 24 70 | Fax: +49 (0)30 22 00 24 799  
info@tmf-ev.de | www.tmf-ev.de

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

© 2016, TMF –Technologie- und Methodenplattform für die vernetzte medizinische Forschung e. V.

Das Werk ist in allen seinen Teilen urheberrechtlich geschützt. Jede Verwertung ohne ausdrückliche Zustimmung des Verlages ist unzulässig. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung in und Verarbeitung durch elektronische Systeme.

Layout: sku:l communication, Michaela Richter, Reichshof-Nosbach

Umsetzung: Anja Brysch, TMF e. V., Berlin

Verlag: Akademische Verlagsgesellschaft AKA GmbH,  
Postfach 22 01 16, 14061 Berlin | [www.aka-verlag.com](http://www.aka-verlag.com)

Druck: Bookstation GmbH, Anzing

Printed in Germany

ISSN: 2363-5754

ISBN: 978-3-89838-723-1

# IT-Reviewing-Board der TMF

**Prof. Dr. Ulrich Sax, Prof. Dr. Otto Rienhoff**

unter Mitwirkung von Dr. Sara Y. Nußbeck, Dr. Nadine Umbach,  
Karoline Buckow, Thomas Franke, Christian Bauer, Dr. Henning Dathe, Thorsten Rottmann  
*Institut für Medizinische Informatik, Universitätsmedizin Göttingen*  
sowie Prof. Dr. Tim Beißbarth,  
*Institut für Medizinische Statistik, Universitätsmedizin Göttingen*

**Prof. Dr. Klaus Pommerening**

*Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsmedizin Mainz*

**Prof. Dr. Frank Ückert**

unter Mitwirkung von Dr. Martin Lablans  
*Medizinische Informatik in der Translationalen Onkologie, Deutsches Krebsforschungszentrum*

**Prof. Dr. Jürgen Stausberg**

*Arzt für Medizinische Informatik und Ärztliches Qualitätsmanagement, Essen*

**Prof. Dr. Christian Ohmann**

*Koordinierungszentrum für Klinische Studien,  
Medizinische Fakultät der Heinrich-Heine-Universität Düsseldorf*

**Prof. Dr. Hans-Ulrich Prokosch**

unter Mitwirkung von Dr. Thomas Ganslandt  
*Lehrstuhl für Medizinische Informatik, Medizinische Fakultät der  
Friedrich-Alexander-Universität Erlangen-Nürnberg*

**Ronald Speer**

unter Mitwirkung von Dr. Toralf Kirsten  
*LIFE Forschungszentrum für Zivilisationserkrankungen, Universität Leipzig*  
sowie Matthias Löbe  
*Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE), Universität Leipzig*

**Prof. Dr. Martin Dugas**

*Institut für Medizinische Informatik, Westfälische Wilhelms-Universität Münster*

**Sebastian C. Semler, Dr. Johannes Drepper**

unter Mitwirkung von Dr. Annette Pollex-Krüger  
*TMF e. V., Geschäftsstelle, Berlin*

# Gastbeiträge

**Dr. Udo Altmann**

*Institut für Medizinische Informatik, Justus-Liebig-Universität Gießen*

**Dr. Fabian Praßer**

*Fakultät für Informatik, Technische Universität München*

**Prof. Dr. Murat Sariyar**

*Institut für Medizininformatik, Berner Fachhochschule*

**Dr. Johannes Janssen**

*Wissenschaftliche Geräte und Informationstechnik, Deutsche Forschungsgemeinschaft (DFG)*

<b>Zusammenfassung</b>	<b>13</b>
------------------------	-----------

---

<b>Einleitung</b>	<b>21</b>
-------------------	-----------

---

<b>Klinische Studien</b>	<b>25</b>
--------------------------	-----------

---

Anwendungsbereich	25
Betreiber, Anwender, Nutzer	26
Rahmenbedingungen	27
Kategorisierung der IT-Landschaft	30
Aktueller Bestand	32
Transparenz bei klinischen Studien	35
Compliance und Validierung computergestützter Systeme in klinischen Studien	38
Bewertung und Handlungsbedarf	39
Literatur	42

<b>Register und Kohorten</b>	<b>47</b>
------------------------------	-----------

---

Anwendungsbereich	47
Rahmenbedingungen	48
Kategorisierung der IT-Landschaft	49
Anforderungen an Kohortenstudien & Register IT (KoRegIT)	50
Aktueller Bestand	55
Verfügbare Softwareprodukte	58
Epidemiologische und klinische Krebsregister	65
Bewertung und Handlungsbedarf	73
Danksagung	76
Literatur	76

<b>Erhebung, Management und Verarbeitung digitaler Bilder</b>	<b>79</b>
---	-----------

---

Anwendungsbereich	79
Kategorisierung der IT-Landschaft	81
Betreiber, Anwender und Nutzer	82

Rahmenbedingungen	83
Aktueller Bestand und Erhebungsergebnisse	85
Bewertung und Handlungsbedarf	91
Literatur	93

## **Biobanken** **95**

---

Anwendungsbereich	95
Betreiber, Anwender, Nutzer	97
Rahmenbedingungen	98
Kategorisierung der IT-Landschaft	101
Aktueller Stand	102
Bewertung und Handlungsbedarf	107
Literatur	110

## **Molekularbiologische Daten aus Hochdurchsatz-Analysen** **115**

---

Anwendungsbereich	115
Betreiber, Anwender, Nutzer	118
Rahmenbedingungen	119
Kategorisierung der IT-Landschaft	129
Aktueller Bestand	130
Bereitstellung und Veröffentlichung von Omics-Daten	133
Bewertung und Handlungsbedarf	136
Danksagung	138
Literatur	138

## **Mobile IT-Werkzeuge** **143**

---

Anwendungsbereich und Eingrenzung	143
Betreiber, Anwender, Nutzer	147
Rahmenbedingungen	147
Kategorisierung der IT-Landschaft	150
Aktueller Stand	152
Bewertung und Handlungsbedarf	158
Literatur	160

## **Sekundärnutzung klinischer Daten** **165**

---

Anwendungsbereich	165
Betreiber, Anwender, Nutzer	167
Rahmenbedingungen	169
Kategorisierung der IT-Landschaft	177
Aktueller Bestand	182
Bewertung und Handlungsbedarf	185
Literatur	188

## **Forschungsdatenmanagement** **195**

---

Anwendungsbereich	195
Kategorisierung der IT-Landschaft	197
Aktueller Bestand und Erhebungsergebnisse	200
Bewertung und Handlungsbedarf	205
Quellen und weiterführende Literatur	206

## **Metadaten und Metadata Repositories** **211**

---

Anwendungsbereich	211
Betreiber, Anwender, Nutzer	214
Rahmenbedingungen	215
Kategorisierung der IT-Landschaft	218
Aktueller Bestand	219
Bewertung und Handlungsbedarf	222
Literatur	224

## **Identitätsmanagement** **229**

---

Anwendungsbereich	229
Betreiber, Anwender, Nutzer	230
Rahmenbedingungen	231
Kategorisierung der IT-Landschaft	233
Aktueller Bestand	234
Bewertung und Handlungsbedarf	236
Literatur	238

## **Anonymisierung von medizinischen Individualdaten 241**

---

Anwendungsbereich	241
Methoden und Anwender	242
Rahmenbedingungen	245
Kategorisierung der IT-Landschaft	246
Aktueller Bestand	247
Bewertung und Handlungsbedarf	249
Literatur	252

## **Cloud-Computing in der medizinischen Forschung 255**

---

Anwendungsbereich	255
Betreiber, Anwender, Nutzer	257
Rahmenbedingungen	258
Aktueller Bestand	265
Bewertung und Handlungsbedarf	267
Literatur	269

## **Strategische Aktivitäten und Fördermaßnahmen im Bereich der IT-Infrastrukturen 275**

---

Strategische Aktivitäten	275
Fördermaßnahmen im Bereich der IT-Infrastrukturen in Deutschland	291
Bewertung und Handlungsbedarf	295
Literatur	297

## **Abkürzungsverzeichnis 301**

---

## Übersicht Infoboxen

---

Study Management Tool des Clinical Trial Centers Aachen	37
IT-Infrastruktur der NAKO Gesundheitsstudie	53
Versorgungsforschung mit und für Menschen mit Multipler Sklerose (MS) – das Deutsche MS-Register	64
Medizinisches Forschungsdatenmanagement an der CAU Kiel	134



# Zusammenfassung

Patientenorientierte medizinische Forschung ist ohne die Unterstützung durch eine geeignete IT-Infrastruktur heute nicht mehr denkbar. Seit über 15 Jahren begleitet die TMF den digitalen Wandel in der medizinischen Forschung, der mit der Digitalisierung der Gesundheitsversorgung in engem Zusammenhang steht. Mit den Berichten des IT-Reviewing-Boards veröffentlicht die TMF seit 2013 eine kontinuierliche Bestands- und Bedarfsanalyse als Grundlage für die strategische Weiterentwicklung der IT-Infrastrukturen in der medizinischen Forschung.

In den ersten beiden Berichten sind die verschiedenen Datenquellen in den Blick genommen worden: von klinischen Studien über Register und Kohorten bis zur Omics-Technologie und der sekundären Nutzung klinischer Behandlungsdaten. Für jeden Bereich wurde dargestellt, welche Lösungen möglich sind bzw. schon eingesetzt werden. Als Querschnittsaufgabe ist bereits das ID-Management beschrieben worden, also die Generierung und Verwaltung eindeutiger Kennzeichner für alle Probanden, Datensätze und Proben. Diese Darstellung hat allerdings in Form einer Leerstelle auf das Problem der Integration dieser heterogenen Daten verwiesen. Wie können die verschiedenen Datenarten zueinander in Beziehung gesetzt und übergreifend ausgewertet werden? Wie eine IT-Architektur hierfür gestaltet sein muss und welche Softwarelösungen notwendig sind, ist entsprechend als Schwerpunktthema im dritten Bericht aufgegriffen worden. Einzelne Aspekte wie Metadatenmanagement, Anonymisierung von Forschungsdaten und Cloud-Computing wurden in spezifischen Kapiteln vertieft behandelt. In der jetzt vorliegenden vierten Ausgabe wurde die bisherige Struktur und Kapitelaufteilung beibehalten. Alle Kapitel wurden umfassend aktualisiert und vermehrt Querbezüge der Kapitel untereinander hergestellt.

Während es für einzelne Datenarten und Forschungsprojekte an ausgewählten Standorten schon sehr gute IT-Lösungen gibt, scheint die Integration immer noch eine große Herausforderung darzustellen. Dies gilt sowohl standortbezogen als auch in Forschungsverbänden und damit standortübergreifend. Auch das BMBF sieht hier Handlungsbedarf und hat genau diese standortbezogene und übergreifende Integration in Form vernetzter Datenintegrationszentren in den Mittelpunkt des Förderkonzepts Medizininformatik gestellt. Vor diesem Hintergrund werden für die weitere Entwicklung der IT-Infrastrukturen in der medizinischen Forschung in Deutschland – teils im Rahmen des Förderkonzepts Medizininformatik des BMBF, teils darüber hinaus – die folgenden Anforderungen und Empfehlungen gesehen:

- ▶ **Sichtbarkeit und Transparenz:** Die Forderung nach einer systematischen Übersicht über bestehende Systeme, Ressourcen und Datenbestände in der medizinischen Forschung zieht sich durch alle im IT-Report angesprochenen Themenfelder. Initiativen hierzu sind in den vergangenen Jahren sowohl aus der TMF-Community hervorgegangen als auch von außerhalb gestartet worden. Allerdings besteht nach wie vor die Schwierigkeit, solche Infrastrukturen aus einer Projektförderung in den langfristigen Betrieb zu überführen. Hier sind geeignete Finanzierungs- und Betriebsmodelle sowie Anreizsysteme zu entwickeln, insbesondere um diese Angebote in einem dynamischen Wissenschaftsumfeld laufend aktuell zu halten. Dies kann nur mit Unterstützung der Politik und der Förderorganisationen geschehen.
- ▶ **Qualitätssicherung und Nachvollziehbarkeit:** In den letzten Jahren ist vermehrt kritisiert worden, dass zu wenig Forschungsergebnisse nachvollziehbar sind, zu wenig experimentell untersuchte Therapieansätze sich in klinischen Studien bewähren oder gar in die Versorgung gebracht werden können und selbst auf Ergebnisse klinischer Studien zu wenig Verlass ist, wenn Studien ohne positive Ergebnisse nicht veröffentlicht werden. Für die heute notwendigen komplexen IT-Infrastrukturen in der Forschung bedeutet dies, dass von sehr hohen Anforderungen an die Qualitätssicherung auf allen Ebenen auszugehen ist. Für jedes Ergebnis muss nachvollziehbar dokumentiert sein, aus welchen Daten es entstanden ist, welchen Weg der Verarbeitung diese Daten genommen haben, welche Nachprüfungen und Veränderungen durch wen vorgenommen, welche Schnittstellen durchlaufen und wie und wann die Daten ursprünglich erhoben wurden. Hierfür muss auf allen Ebenen und immer wieder das Bewusstsein hoch gehalten oder auch erst geschaffen werden.
- ▶ **Standardisierung und Interoperabilität:** Medizinischer Fortschritt beruht auf der Möglichkeit, Informationen aus verschiedenen Sektoren und IT-Systemen zusammenzuführen. Neben der Harmonisierung der Daten- und Übertragungsstandards ist eine standardisierte semantische Annotation der Daten der beteiligten Systeme notwendig. Hierfür müssen vorhandene und z.T. kostenfrei nutzbare Standards mehr als bisher eingesetzt werden. Darüber hinaus ist aber auch ein systematisch begleiteter Einstieg in die Nutzung der Referenz-Terminologie SNOMED-CT notwendig. Zudem können Datensammlungen erst dann übergreifend genutzt werden, wenn auch die Metadaten standardisiert erfasst und automatisiert auswertbar sind. Diese tiefgreifenden und die Einbindung vieler Parteien aus Forschung und Versorgung erfordernden Maßnahmen können nicht ohne den Aufbau einer geeigneten nationalen Steuer- und Begleitstruktur gelingen. Parallel ist die Förderung der Forschung und des Nachwuchses auf dem Feld der medizinischen Terminologie auszubauen.

- ▶ **Erfahrungsaustausch und Abstimmung:** Analog zu Daten und Ressourcen gilt es auch, das Wissen und die Erfahrungen mit dem Aufbau und Betrieb von IT-Infrastrukturen für die patientenorientierte Forschung zusammenzuführen, auszutauschen, gemeinsam weiterzuentwickeln und für nachfolgende Projekte verfügbar zu machen. So können IT-Konzepte und -Aufbauerfahrungen aus Leuchtturmprojekten und Vorreiterstandorten auch auf andere Einrichtungen ausstrahlen und den Forschungsstandort Deutschland insgesamt voranbringen. Eine unkoordinierte Implementierung von IT-Systemen dagegen führt zu Ineffizienzen und behindert die Weiterentwicklung von Standardisierung und Interoperabilität.
- ▶ **Internationale Anschlussfähigkeit:** Medizinische Forschung geschieht heute sowohl eingebettet in einen weltweiten Informationsaustausch als auch im internationalen Wettbewerb stehend. Gerade die dynamische Methodenentwicklung im Bereich der Omics-Technologien zeigt, dass deutsche Forschungsstandorte von internationalen Tools, Methoden und Standards nicht nur profitieren können, sondern auch müssen. Unnötige Eigenentwicklungen und unabgestimmte Alleingänge im methodischen Bereich sind unökonomisch und führen schlimmstenfalls zu qualitativ nicht ausreichend abgesicherter Forschung sowie fehlender internationaler Anschlussfähigkeit.
- ▶ **Aufbau von Dienstleistungsangeboten:** Vermehrt wird der Bedarf geäußert, im Bereich der akademischen medizinischen Forschung zentral angebotene Dienstleistungen zu moderaten Preisen zu etablieren. Dies gilt von der Nutzung einzelner Softwarekomponenten über das Angebot eines Identitätsmanagements im Rahmen der benötigten Datenschutzkonzepte bis hin zu einer deutschen Cloud für „Applied & Translational Genomics“. Bedarfsermittlung, Planung und Aufbau solcher notwendigerweise standardisierten Services können nur in Standort-übergreifender Abstimmung geschehen. Notwendig für die Koordination ist eine neutrale und außerhalb des wissenschaftlichen Wettbewerbs stehende Stelle.
- ▶ **Klärung rechtlicher Fragen:** Mit der Weiterentwicklung des medizinischen Wissens und der technologischen Möglichkeiten sind auch weiterhin neue rechtliche und ethische Fragen zu beantworten. Das gilt unter anderem für den zunehmenden Einsatz mobiler Technologien in medizinischer Forschung und Versorgung, im Bereich der Sekundärnutzung von Versorgungsdaten für Forschung und Qualitätssicherung sowie nicht zuletzt auch für den Umgang mit genetischen Daten aus Hochdurchsatz-Analysen (Omics-Technologien).

- ▶ **Finanzierung:** IT-Infrastrukturen in der medizinischen Forschung sind nach wie vor nicht ausreichend finanziert. IT-Systeme werden häufig im Rahmen von zeitlich begrenzten (Verbund-)Projekten aufgebaut, weshalb sie oft zu wenig in die Gesamtstruktur einer Einrichtung eingebettet sind und nach Ablauf des Förderzeitraumes nicht mehr optimal betrieben werden können. Besonders problematisch ist auch, dass qualifiziertes Personal, das über das Wissen und die Erfahrung im Umgang mit den IT-Systemen verfügt, angesichts von Zeitverträgen im Rahmen von Projekten oftmals nicht gehalten werden kann. Das Förderkonzept Medizin-informatik des BMBF wird vor diesem Hintergrund als positiver Impuls gewertet. Diese Fördermaßnahme fokussiert sinnvollerweise auf die Universitätskliniken als zentrale Einrichtungen der patientenorientierten Forschung. Gerade für IT-Infrastrukturen dieser Einrichtungen müssen aber gemeinsam zwischen Bund und Ländern abgestimmte sowie langfristig wirksame und auch Betriebskosten umfassende Finanzierungsstrategien gefunden werden.

## Entwicklungen und Fortschritte seit 2015

- ▶ Im April 2016 wurde nach über vierjähriger Verhandlung die neue EU-Datenschutzgrundverordnung verabschiedet. Diese wird ab Mai 2018 direkt geltendes Recht in allen EU-Staaten sein. Das Ziel einer EU-weiten Harmonisierung des Datenschutzrechts ist auch ein wichtiges Anliegen der biomedizinischen Forschung. Europäische Projekte, die Forschung auf Basis von Patientendaten betreiben, leiden heute unter den unterschiedlichen Rechtsauffassungen und Rechtspraxen in den beteiligten Ländern. Gleichwohl lässt die neue Verordnung gerade für die Verarbeitung von Gesundheitsdaten den Nationalstaaten große Spielräume, so dass noch unklar ist, wie weitgehend das Harmonisierungsversprechen tatsächlich in Erfüllung gehen wird. Für den deutschen Forschungsraum steht sogar zu befürchten, dass die Heterogenität der Landesdatenschutzgesetze und weiterer relevanter spezialgesetzlichen Regelungen auf Landesebene bestehen bleiben könnte. Hier bleibt abzuwarten, wie die Weichen auf Bundes- und Landesebene im Rahmen des notwendigen legislativen Anpassungsprozesses gestellt werden.
- ▶ Gerade angesichts der manchmal unklaren oder heterogenen Rechtslage im Datenschutz versuchen Forscher dieser durch die Nutzung anonymisierter Daten zu entkommen. Hier wurde von verschiedenen Seiten schon länger darauf hingewiesen, dass auch die Einschätzung dessen, was für Datensätze tatsächlich als anonym gelten können, üblicherweise von Experte zu Experte differiert. Klar scheint heute hingegen, dass im Regelfall die reine Entfernung klar identifizierender Daten wie etwa Namen, Geburtsdatum und Adresse nicht hinreichend ist, da auch die medizinischen Daten in ihrer heute üblichen Detailliertheit ein zu hohes Reidentifizierungs-

potential bieten. Die TMF hat 2016 im Rahmen eines Projekts ein ausführliches Schulungsprogramm dazu entwickelt, wie man mit heutigen Tools das Reidentifizierungsrisiko z. B. entlang des mathematischen Modells der  $k$ -Anonymität quantifizieren und absenken kann. Die drei im Rahmen der Entwicklung durchgeführten Schulungen waren deutlich übernachgefragt, so dass hier die Fortführung bzw. der Ausbau entsprechender Angebote empfehlenswert erscheint.

- ▶ Wie man hingegen mit pseudonymen Datensammlungen in der medizinischen Forschung umgeht, ist in bewährter und mit den Datenschutzaufsichtsbehörden auf nationaler Ebene abgestimmter Form im Datenschutzleitfaden der TMF von 2014 beschrieben worden. Allerdings werden von den heute bekannten Pseudonymisierungstools noch nicht alle im Leitfaden beschriebenen Pseudonymisierungsvorgänge unterstützt. Zudem lassen sich die an verschiedenen Standorten entwickelten Werkzeuge bisher mangels einheitlicher und standardisierter Schnittstellen kaum kombiniert einsetzen. 2016 konnte nun das Projekt MAGIC im Rahmen des Förderprogramms der DFG für Forschungsprojekte über die und mit der TMF gestartet werden, welches genau diese Abstimmung und Ergänzung bestehender Tools zum Ziel hat und somit für eine breit einsetzbare Referenzimplementierung aller im Leitfaden beschriebenen Verfahren sorgen kann.
- ▶ Neben weiteren wichtigen Projekten aus dieser DFG-Fördermaßnahme ist insbesondere das Projekt zum Aufbau eines Community-basierten nationalen Metadaten Repositories zu nennen, welches Anfang 2017 die Arbeit aufnehmen wird. Die Erkenntnis, dass in Forschungsprojekten effektiver mit Metadaten umgegangen werden muss, dass diese standardisiert erfasst und automatisiert auswertbar sein müssen, so dass übergreifende Auswertungen möglich werden und Forschung insgesamt nachvollziehbarer wird, zieht sich durch fast alle Kapitel dieses Reports. Mit diesem Projekt besteht die Hoffnung, dass alle relevanten Stakeholder und Know-how-Träger zu diesem Themenfeld, nicht zuletzt auch mit der Hilfe der an dem Projekt beteiligten TMF, zusammengebracht werden können und eine für alle Forschungsstandorte hilfreiche Infrastruktur entsteht.
- ▶ Die TMF hat in den letzten Jahren im Rahmen des vom BMBF geförderten Projekts MethInfraNet (Maßnahmen zur methodischen und infrastrukturellen Vernetzung für Qualitäts- und Effizienzsteigerung in der medizinischen Forschung) das Web-Portal „ToolPool Gesundheitsforschung“ zur Unterstützung des Aufbaus und Betriebs von IT-Infrastrukturen in der klinischen Forschung konzipiert und entwickelt. Das Ziel ist, die IT-Infrastrukturen, die für medizinische Forschung benötigt werden, und damit in Zusammenhang stehende Themen stärker ins Bewusstsein der Akteure (Forscher, Medizininformatiker, Wissenschaftsmanager, Förderorganisationen) zu rücken und konkrete Anleitung und Unterstützung anzubieten. Das von der Community der TMF getragene und unterstützte Portal bündelt Produkte verschiedenster Anbieter

und bietet leicht verständliche Einführungen in die beim Aufbau und Betrieb von IT-Infrastrukturen in der medizinischen Forschung entstehenden Anforderungen. Die thematische Struktur des Portals ist an die Struktur des vorliegenden IT-Reports angelehnt, so dass eine enge inhaltliche Verzahnung dieser beiden Angebote erreicht werden kann. Nach einer umfangreichen Testphase 2016 ist der Start des Betriebs für Anfang 2017 geplant.

- ▶ Beim Einsatz von IT-Systemen in der klinischen Forschung ist eine zentrale Anforderung des Qualitätsmanagements die Validierung derartiger Systeme. 2016 wurde in einem TMF-Projekt die hierfür aus einer früheren Initiative bestehende Sammlung an Musterdokumenten, Checklisten und Vorlagen überarbeitet. Zudem wurde ein bestehendes Schulungskonzept angepasst und anhand einer beispielhaften Schulung evaluiert. Ergänzend wird ein Auditierungskonzept erarbeitet und abgestimmt, welches Grundlage eines gegenseitigen Auditierens der akademischen Standorte werden soll. Aufbauend auf diesen Konzepten und Aktivitäten wurde ein Dialog mit den Aufsichtsbehörden gestartet, der langfristig auch in konsentrierte Verfahren münden soll.
- ▶ Das sichere und effiziente Bereitstellen von IT-Anwendungen, wie es auch für die biomedizinische und klinische Forschung unverzichtbar ist, erfordert angesichts steigender Anforderungen an Funktionalität, komplexer werdender Basistechnologie und einem heute schon schwer einschätzbaren Sicherheitsumfeld ein hohes Ausbildungs- und Kenntnisniveau in den IT-Abteilungen. Neben einer immer wieder nötigen Prüfung einer neuen Aufgabenverteilung erfordert dies auch eine sehr effiziente Einarbeitung in und Adaptation neuer Technologien, wie sie gerade auch im Umfeld des Cloud-Computings entstehen. Vor diesem Hintergrund ist die in 2016 erfolgte umfangreiche Auseinandersetzung mit den Nutzungsmöglichkeiten von Docker-Containern in der medizinischen Forschung im Rahmen einer TMF-Projektserie mit der Bereitstellung diverser Tools in Containern und der Durchführung zweier öffentlicher Workshops positiv zu werten.
- ▶ Der Trend zur vermehrten Nutzung von Clouds betrifft auch die IT-Infrastrukturen in der medizinischen Forschung. Eine besondere Dynamik bekommt diese Entwicklung durch die Forderung zur Bereitstellung von immer mehr Forschungsdaten sowohl zur Überprüfung bisher publizierter Forschungsarbeiten als auch der Generierung neuer Ergebnisse. Die Europäische Kommission sieht entsprechenden Handlungsbedarf und hat 2016 eine Experten-Gruppe ein erstes Konzeptpapier für eine European Open Science Cloud (EOSC) erarbeiten lassen. Darin wird insbesondere auf die Notwendigkeit der sorgfältigen Aufbereitung, Bewahrung und Verfügbarmachung der enormen Datenmengen in der heutigen Wissenschaft hingewiesen. Hierzu wird ein umfangreiches Förderprogramm für entsprechende Cloud-Infrastrukturen in der EU vorgeschlagen. Wichtig ist, dass im Zuge einer

solchen Fördermaßnahme auch technische und organisatorische Konzepte entwickelt werden, die später die Nutzung einer solchen verteilten Cloud-Infrastruktur für sensible Gesundheitsdaten ermöglichen.

- ▶ Ende 2015 erfolgte die Ausschreibung des BMBF zur „Ertüchtigung deutscher Biobanken für die Anbindung an BBMRI“, mit der die sogenannte German Biobank Alliance (GBA) etabliert werden soll. In diesem Rahmen sollen von 2017 bis 2019 zehn Biobankstandorte unter der Koordination des German Biobank Node (GBN) u. a. IT-Standards und Werkzeuge für die deutsche Biobanking Community entwickeln und testen. Es ist zu hoffen, dass diese Initiative den Professionalisierungstrend im Biobanking in Deutschland weiter voran bringt und durch entsprechende Begleitaktivitäten sichergestellt wird, dass die zunehmende Qualität der Lagerung und Verwaltung der Bioproben sowie die weitergehende Vernetzung auf nationaler und europäischer Ebene nicht auf die zehn geförderten Standorte beschränkt bleibt.
- ▶ Einen nachhaltigen Einfluss auf die Entwicklung der Medizininformatik in Deutschland wird das Förderkonzept „Medizininformatik“ des BMBF haben, welches im Sommer 2016 in die Konzeptphase gestartet ist. Explizites Ziel dieser Initiative ist es, Austausch und Nutzung von Daten nicht nur zwischen Institutionen, sondern insbesondere auch zwischen der Krankenversorgung und der klinischen bzw. biomedizinischen Forschung zu intensivieren. Hierfür arbeiten derzeit sieben Konsortien, in denen fast alle Universitätsklinika in Deutschland vertreten sind, an der Konzeption einrichtungsübergreifend anschlussfähiger Datenintegrationszentren, die sowohl Versorgungs- als auch Forschungsdaten in standardisierter Form zugänglich machen sollen. Die hierfür nötigen technischen, organisatorischen und rechtlichen Rahmenbedingungen werden in einer nationalen Begleitstruktur erarbeitet und abgestimmt. Wichtig ist, dass dies auf der Ebene der Begleitstruktur verzahnt mit anderen großen Förderinitiativen wie den Deutschen Zentren der Gesundheitsforschung, der NAKO Gesundheitsstudie oder auch der German Biobank Alliance geschieht. Es ist zu hoffen, dass diese wichtige Förderinitiative dazu genutzt werden kann, zu vielen der aufgezeigten Handlungsfelder in den nächsten Jahren deutliche Fortschritte zu verzeichnen.



# Einleitung

Wer IT-Infrastrukturen in der Forschung auf- oder ausbauen möchte, steht vor einer großen Zahl von Fragen und Entscheidungen, z. B.: Was muss ich bei einer Beschaffung von Software für klinische Studien zu neuen Arzneimitteln berücksichtigen? Mit welchen Betriebs- oder Folgekosten muss ich rechnen? Welche Anforderungen sind bei der Auswahl einer Software für Register zu beachten? Welche Tools unterstützen eine Pseudonymisierung der Daten? Und wenn ich genetische Daten, Bilder oder Biomaterialien zusätzlich verwalten will – wo bekomme ich Informationen zu passender Software und wie kann ich diese an andere IT-Infrastrukturen anbinden? Kann ich auch die Daten aus einem Klinikinformationssystem für meine Forschungsprojekte nutzen? Wie kann ich all diese Daten aus unterschiedlichen Quellen für neue Forschungsprojekte aufbereiten und nutzen bzw. nachnutzen?

Auch diejenigen Forscher, die nicht direkt für die IT zuständig sind, wie z. B. Instituts- oder Projektleiter, Forschungskoordinatoren oder Antragsteller, müssen sich eine Reihe von Fragen stellen: Für welche Aufgaben gibt es bereits IT-Lösungen und ggf. Kooperationspartner mit umfangreichen Erfahrungen? Wie müssen die Forschungsfragestellungen an die gesetzlichen Rahmenbedingungen angepasst werden? Welche Finanzmittel werden für die IT in den nächsten Jahren oder in dem zu beantragenden Projekt benötigt? Und kann die im Projekt aufgebaute IT-Infrastruktur nach Auslaufen der Förderung weiter betrieben werden?

Und auch von Mitarbeitern von Forschungsförderorganisationen oder Gutachtern wird immer mehr IT-Expertise erwartet. Sie haben Fragen zu beantworten wie etwa: In welchem Bereich ist in Deutschland ausreichend methodische Kompetenz vorhanden und wo ist ggf. eine Ausschreibung für Methodenprojekte sinnvoll? Welche Anteile einer IT-Infrastruktur müssen in einem Forschungsprojekt mit gefördert werden? Auf welche Hilfsmittel können Antragsteller hingewiesen werden? Was ist State of the Art?

Um diese hier nur beispielhaft genannten und viele weitere Fragen zur Forschungs-IT verlässlich beantworten zu können, ist ein regelmäßiger Austausch der Beteiligten erforderlich, wie er zum Beispiel in den Arbeitsgruppen der TMF organisiert wird. Gleichzeitig wird als unterstützende Maßnahme eine kontinuierliche Bestands- und Bedarfsanalyse benötigt, die zum einen die Ergebnisse des Austauschs in konsolidierter Form bereitstellen und zum anderen auch auf Wissenslücken und Handlungsbedarf hinweisen kann.

Für die Konzeption und Durchführung einer solchen Bestandsaufnahme von IT-Strukturen wurde mit Unterstützung des Bundesministeriums für Bildung und Forschung (BMBF) ein IT-Reviewing-Board gegründet, welches die Ergebnisse in Form eines öffentlichen Berichts zur Verfügung stellt. Das Board setzt sich aus Experten für IT-Infrastrukturen in der medizinischen Forschung und Versorgung zusammen. Die Mitarbeit in dem Board setzt neben der notwendigen Expertise die Bereitschaft voraus, sich für einen längeren Zeitraum zu einem relevanten Themenbereich zu engagieren und an der Darstellung des Themenbereichs in den regelmäßig aktualisierten Berichten zu beteiligen. Von den aktuellen Mitgliedern des IT-Reviewing-Boards sind die meisten seit vielen Jahren in Gremien, Projekten oder anderen Querschnittsaktivitäten der TMF aktiv und haben die Arbeit der TMF zum Teil entscheidend mit geprägt. Die Mitwirkung im Board ist allerdings nicht auf den Kreis der TMF-Mitglieder beschränkt.

Das Board ist bei der TMF als Dachorganisation der medizinischen Verbundforschung in Deutschland angesiedelt. Die TMF ist eine Plattform für den interdisziplinären Austausch und die projekt- wie standortübergreifende Zusammenarbeit, um gemeinsam die organisatorischen, rechtlich-ethischen und technologischen Probleme der modernen medizinischen Forschung zu identifizieren und zu lösen. Die TMF bringt Forscher unterschiedlicher Disziplinen zusammen und entwickelt Konzepte, Methoden, Infrastrukturen und Lösungen.

Der Report des IT-Reviewing-Boards wird zurzeit jährlich mit wechselnden Schwerpunkten, Themen und Autoren fortgeschrieben. Schwerpunkt der ersten Ausgabe war die Datenerhebung. Die zweite Ausgabe schrieb diesen Schwerpunkt fort, enthielt jedoch erstmals auch jeweils ein Kapitel zur Erhebung und Verarbeitung molekularbiologischer Daten sowie zur Sekundärnutzung klinischer Daten für die Forschung. In der dritten Ausgabe wurde der Fokus auf Querschnittsaspekte gelegt, beispielsweise auf die Erfassung und Verwaltung von Metadaten sowie insbesondere auf die Aufgaben des Forschungsdatenmanagements (vgl. Abb. 1.1). Einzelne Aspekte des Forschungsdatenmanagements, zu denen es in der Vergangenheit in den Gremien und Projekten der TMF viele Fragen gegeben hat, wurden zusätzlich in separaten Kapiteln behandelt. Hierzu zählt die Anonymisierung von Forschungsdaten und das Auslagern von Forschungsdaten und Verarbeitungsprozessen in die Cloud. In der jetzt vorliegenden vierten Ausgabe wurde die bisherige Struktur und Kapitelaufteilung beibehalten. Allerdings wurden alle Kapitel umfassend aktualisiert und vermehrt Querbezüge der Kapitel untereinander hergestellt.

Die Kapitel im vorliegenden Bericht sind im Regelfall in jeweils sechs Abschnitte unterteilt. Dargestellt werden der Anwendungsbereich, die handelnden Akteure und ihre Rollen, relevante Rahmenbedingungen, die Kategorisierung der IT-Landschaft und der aktuelle Bestand. Abschließend folgt eine Bewertung mit dem sich daraus eventuell ergebenden Handlungsbedarf.

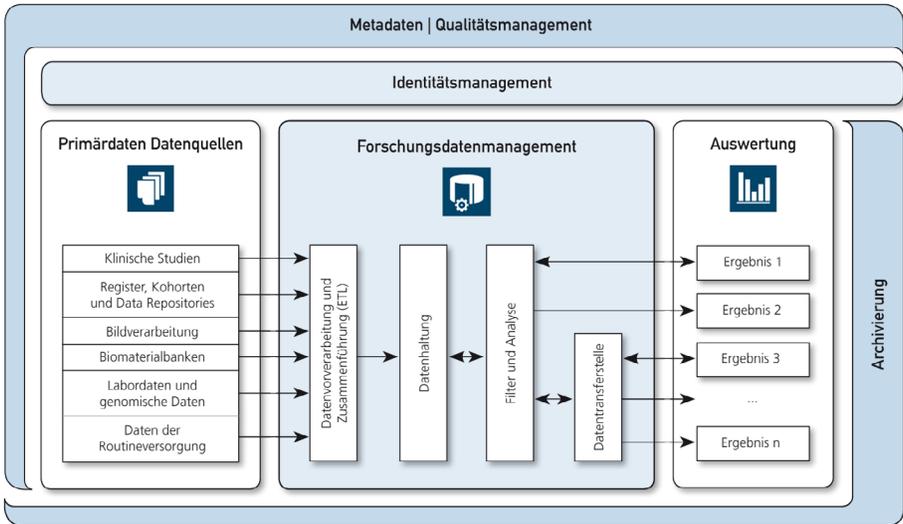


Abbildung 1.1: Generische Darstellung der Datenverarbeitung in der patientenorientierten Forschung von den Datenquellen bis zur Auswertung und Archivierung. Neben einer ausführlichen Darstellung der Datenquellen wird auch auf verschiedene Querschnittsaspekte und insbesondere die Aufgaben des Forschungsdatenmanagements eingegangen. Hierzu gehört das selektive Lesen von Daten aus mehreren, heterogenen Datenquellen (Extract), das Transformieren (Transform) und die Abspeicherung in einer einheitlichen Zielstruktur (Load) – zusammen oft als ETL-Prozess abgekürzt bezeichnet, aber auch die Datenhaltung, der Schritt der Filterung und Analyse von Daten sowie die Unterstützung von und Zusammenarbeit mit Datentransferstellen wird beleuchtet.

Auch der vorliegende Bericht schließt – wie bereits der vorhergehende – mit einem Kapitel zu strategischen Aktivitäten und Fördermaßnahmen im Bereich der IT-Infrastrukturen. Hier sind in einzigartig konzentrierter Form die zahlreichen Initiativen, Gremien, Stellungnahmen, Empfehlungen, Leitlinien und Roadmaps zu Forschungsinfrastrukturen mit IT-Bezug nebeneinander dargestellt. So werden Überlappungen, unterschiedliche Akzent- und Schwerpunktsetzungen sowie alternative Lösungsansätze erkennbar. Zusammen mit einer Übersicht der Fördermaßnahmen in diesem Bereich macht dieses Kapitel die Vielfältigkeit deutlich, mit der auf nationaler wie europäischer Ebene das zentrale Thema der IT-Infrastrukturen in der medizinischen Forschung behandelt wird.

## Danksagung

Ein Großteil der in diesem Bericht zusammengetragenen Informationen geht auf den regelmäßigen interdisziplinären Austausch in den Arbeitsgruppen der TMF zurück, die ihre Sitzungen in den letzten Jahren auch verstärkt für Teilnehmer und Anregungen von außerhalb der Mitgliedschaft geöffnet haben. Dies sind neben den Arbeitsgruppen „IT-Infrastruktur und Qualitätsmanagement“, „Biomaterialbanken“ und „Datenschutz“ auch die Arbeitsgruppen „Management klinischer Studien“ und „Molekulare Medizin“. Weitergehende Informationen zu den Arbeitsgruppen und den Möglichkeiten, als Mitgliedsvertreter oder Gast an diesem Austausch teilzunehmen, vermitteln gerne die jeweiligen Ansprechpartner der Geschäftsstelle der TMF.<sup>1</sup>

Dieses Projekt zur Erstellung eines IT-Reviews war nur möglich durch die Förderung des Projekts „Methodische und infrastrukturelle Vernetzung für Qualitäts- und Effizienzsteigerung in der medizinischen Forschung“ (MethInfraNet) durch das Bundesministerium für Bildung und Forschung (Förderkennzeichen: 01GI1003). Wir bedanken uns an dieser Stelle bei den Verantwortlichen im BMBF für die Weitsicht und die finanzielle Förderung.

---

<sup>1</sup> Siehe [www.tmf-ev.de](http://www.tmf-ev.de).

# Klinische Studien

*Christian Ohmann, Ronald Speer*

*Mitarbeit: Thomas M. Deserno<sup>1</sup>*

## Anwendungsbereich

In diesem Abschnitt wird die Erfassung und Verarbeitung von Daten im Rahmen klinischer Studien behandelt. Die Organisation und Durchführung von klinischen Studien ist eines der zentralen Themen im Bereich der klinischen Forschung und ohne den Einsatz von rechnergestützten Verfahren nicht mehr vorstellbar. Erfassung und Verarbeitung der Daten in klinischen Studien stellen dabei große Anforderungen an die unterschiedlichen Werkzeuge und Softwarelösungen, die hierfür eingesetzt werden. Diese unterscheiden sich je nach Anforderungsbereich in ihrer Funktionalität. Für die Erfassung und Verwaltung der Daten werden Clinical Data Management Systems (CDMS) genutzt. Die Planung und Verwaltung von Studien wird hingegen unter anderem durch Software zum Projekt- und Dokumentenmanagement unterstützt. Spezialsoftware kommt z. B. für die Erfassung, Verwaltung und Meldung von unerwünschten Ereignissen (Serious Adverse Events, SAEs) zum Einsatz. Weiterhin werden Softwarelösungen für die Verwaltung von Studienteilnehmern, Terminmanagement und die Erfassung des studienbezogenen Mehraufwands eingesetzt. Für die Auswertung der Daten werden zudem verschiedene Statistikpakete eingesetzt (z. B. SAS, R, SPSS). Im Folgenden werden vorrangig Softwarelösungen zur Erfassung und Verwaltung der Daten in klinischen Studien behandelt. Ebenfalls werden die Anforderungen sowie vorhandene Lösungen für das Management von Studienteilnehmern und Untersuchungsterminen betrachtet.

Generell bestehen für alle Aufgaben im Bereich klinischer Studien hohe Anforderungen an das Qualitätsmanagement. Zudem existiert in diesem Bereich eine Vielzahl an Gesetzen und Richtlinien, die zu beachten sind.

Fazit: Die Erfassung und Verwaltung von Daten in klinischen Studien ist im Bereich der klinischen Forschung eine Kernaufgabe. Die Komplexität der Werkzeuge, die hohen Anforderungen an das Qualitätsmanagement sowie die Vielzahl zu berücksichtigender Regularien stellen dabei die Forschungsverbünde und akademischen Studienzentren vor große Herausforderungen. Das KKS-Netzwerk und die TMF haben hier bereits große Vorarbeiten zur Unterstützung der Verbünde bei der Bewältigung dieser Aufgaben geleistet.

---

<sup>1</sup> Thomas M. Deserno hat die Infobox zum Study Management Tool des Clinical Trial Centers Aachen beigesteuert.

## Betreiber, Anwender, Nutzer

Der Einsatz von Studiensoftware kann in Studienzentren auf der Basis eigener Serverinstallationen oder auch im Rahmen eines Application Service Providing (ASP) erfolgen. Bisher sind die akademischen Studienzentren, wie z. B. die Koordinierungszentren für Klinische Studien (KKS) bzw. die Zentren für Klinische Studien (ZKS), in der überwiegenden Zahl sowohl Betreiber als auch Anwender von Softwarelösungen zur Unterstützung klinischer Studien. Dem gegenüber stehen kommerzielle Organisationen, wie z. B. Contract Research Organisations (CROs) oder pharmazeutische Unternehmen bzw. Unternehmen der Medizintechnik, die vornehmlich für Zulassungsstudien Softwarelösungen betreiben und anwenden oder diese outsourcen. Darüber hinaus betreiben verschiedene Kliniken sowohl im privaten als auch im öffentlichen Sektor eigene Software im Bereich der klinischen Studien.

In den letzten Jahren ist das Interesse an Cloud-Computing zur Unterstützung des Datenmanagements in klinischen Studien erheblich gestiegen. Gründe hierfür sind die Aussicht auf eine als Service verfügbare, preiswerte und flexible IT-Infrastruktur und eine Reduzierung der Anforderungen an Inhouse-Systeme und Personal. Konsequenterweise gibt es immer mehr CDMS-Anbieter, die ihre Systeme als Software as a Service (SaaS) und über cloud-basierte Datenzentren anbieten (z. B. OpenClinica, Medidata, XClinical). In einem ECRIN-Workshop mit internationalen Experten am 30.10.2014 in Brüssel wurden Nutzen und Risiken cloud-basierter Services für klinische Studien diskutiert und Ansatzpunkte für das weitere Vorgehen identifiziert (z. B. Klärung studienspezifischer Anforderungen, Einbindung von Behörden) [1]. Von der Working Group on Cloud Adoption der Pharmaceutical User Software Exchange (PhUSE) wurde ein Framework zu Cloud Services in den regulierten Life Sciences entwickelt, welches derzeit zur Konsultierung ins Netz gestellt wurde [2]. Die europäische Zulassungsbehörde EMA (European Medicines Agency) hat sich ebenfalls dieses Themas angenommen und im Rahmen eines Meetings der GCP Inspectors Working Group (WG) am 30.11.2015 in London auf der Basis dieses Frameworks die Diskussion zu Cloud Services intensiviert. Es ist zu erwarten, dass in absehbarer Zeit Stellungnahmen und Empfehlungen zur Benutzung von Cloud Services bei klinischen Studien vorliegen werden.<sup>2</sup>

Weiterhin führt die Verwaltung der Studienteilnehmer und der betreffenden Einwilligungserklärungen zu einem hohen Aufwand. In diesem Zusammenhang ist auch das Management von Visiten und die Verwaltung zusätzlicher Daten (Biomaterialdaten

---

2 Siehe hierzu auch das Kapitel „Cloud-Computing in der medizinischen Forschung“ im vorliegenden Report.

etc.) zu sehen. Eine Bestandsaufnahme der TMF im Rahmen eines Projektes zeigte, dass hier sehr heterogene Systeme zum Einsatz kommen.<sup>3</sup> Die Gründe hierfür sind in der Regel datenschutzrechtliche Rahmenbedingungen oder die Anforderung der Einbindung lokaler Identitätsmanagementsysteme. Allerdings existieren auch zentral verwaltete Lösungen für das Management von Identifikationsdaten und Visiten von Studienteilnehmern.

Bei den Nutzern kann man folgende Rollen unterscheiden: Sponsoren, Prüfarzte und Mitglieder der Prüfgruppe sowie Personal der Studienzentren (z. B. Projekt- und Datenmanager, Monitore). Bei Studien nach Arzneimittelgesetz (AMG) oder Medizinproduktegesetz (MPG) kommt dem Sponsor eine zentrale Rolle zu, da er für die Planung, Durchführung und Auswertung einer klinischen Studie die Gesamtverantwortung trägt. Prüfarzte und Mitglieder der Prüfgruppe nutzen Studiensoftware zur Eingabe von klinischen Daten. Das Personal von Studienzentren (z. B. KKS, ZKS) ist für die Qualitätssicherung der Daten zuständig und führt bei zentraler Datenerfassung auch die Dateneingabe durch.

## Rahmenbedingungen

Bei dem Einsatz von Software in klinischen Studien kommt eine Reihe rechtlicher Bestimmungen zum Tragen. Wesentliche Grundlagen der klinischen Studienforschung sind in Deutschland das AMG für Studien mit Arzneimitteln und das MPG für Studien mit Medizinprodukten. In beiden Bereichen stehen allerdings Änderungen vor der Tür: Am 16. Juni 2014 ist bereits die neue EU-Verordnung Nr. 536/2014 des Europäischen Parlaments und des Rates über klinische Prüfungen mit Humanarzneimitteln und zur Aufhebung der Richtlinie 2001/20/EG in Kraft getreten. Sie wird – in Abhängigkeit von den bis dahin zu schaffenden Rahmenbedingungen – frühestens ab Oktober 2018 gelten.<sup>4</sup> Im November 2016 hat der Bundestag nach langer Diskussion, die insbesondere die Regelungen zur Forschung mit Nicht-Einwilligungsfähigen Patienten betrafen, die Reform des nationalen Arzneimittelrechts beschlossen. Teile davon sollen Anfang 2017 in Kraft treten, Änderungen zum Genehmigungsverfahren für klinische Prüfungen mit Humanarzneimitteln werden voraussichtlich erst im zweiten Halbjahr 2018 wirksam.<sup>5</sup> Seit Oktober 2012 wird zudem über einen neuen europäischen Rechtsrahmen für Medizinprodukte und In-vitro-Diagnostika im Europäischen Rat und im Europäischen Parlament verhandelt. Im Mai 2016 wurde in den zuständigen europäischen Gremien nach zähen Verhandlungen eine finale Einigung erzielt, so dass noch 2016 mit einer

---

3 <http://www.tmf-ev.de/news/1334>

4 Siehe [http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general\\_content\\_000629.jsp](http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000629.jsp).

5 Siehe <http://www.bundesgesundheitsministerium.de/ministerium/meldungen/2016/4-amg-novelle-verabschiedet.html>.

Verabschiedung und Veröffentlichung gerechnet werden kann. Die neuen Regelungen wären dann ab Ende 2019 anzuwenden.

Von herausragender Bedeutung für die Durchführung von klinischen Studien sind die international anerkannten, nach ethischen und wissenschaftlichen Gesichtspunkten aufgestellten Regeln zur Good Clinical Practice (GCP) [3]. Dabei stehen der Schutz der Studienteilnehmer und deren informierte Einwilligung sowie die Sicherstellung der Qualität der Studienergebnisse im Mittelpunkt. In den GCP-Regeln sind unter anderem Anforderungen definiert, die spezifische Vorgaben für den Einsatz von Studiensoftware in klinischen Studien machen. Da in der Regel in den Studienprojekten personenbezogene Daten erfasst und verarbeitet werden, fallen diese Prozesse je nach Trägerschaft der beteiligten Einrichtungen auch unter das Bundesdatenschutzgesetz (BDSG) oder die Landesdatenschutzgesetze der beteiligten Bundesländer. Diese Gesetze regeln die grundsätzlichen gesetzlichen Anforderungen bei der Erfassung, Verarbeitung und Weitergabe personenbezogener Daten im Rahmen klinischer Studien. Zusätzlich gelten je nach Studientyp und Studienprotokoll weitere Gesetze und Regelungen (z. B. die Strahlenschutzverordnung).

Das EU-geförderte Projekt ECRIN (European Clinical Research Infrastructures Network) hat in einem Konsensusprozess Requirements für Datenzentren, die klinische Studien durchführen, entwickelt und publiziert [4, 5]. Derzeit gültig ist die Version 3.0 [6]. Es handelt sich dabei um einen offenen Standard zur Sicherstellung eines GCP-konformen Datenmanagements in klinischen Studien, der sich vornehmlich an akademische Studienzentren wendet. Auf der Basis des ECRIN-Standards wurde ein Zertifizierungsverfahren für Datenzentren aufgebaut, welches im Rahmen eines EU-geförderten Pilotprojekts erfolgreich implementiert werden konnte. Mittlerweile wurde mit ECRIN-ERIC eine europäische Infrastruktur mit legalem Status aufgebaut. ECRIN-ERIC hat die Zertifizierung von Datenzentren mit regelmäßigen jährlichen Ausschreibungen in einen dreijährigen Arbeitsplan (2015-2017) aufgenommen. Die letzte Ausschreibung erfolgte im Mai 2016 und befindet sich derzeit in der Bearbeitung. Details zu der Zertifizierungsprozedur sind der ECRIN-Webpage zu entnehmen.<sup>6</sup> Insgesamt konnten bisher 9 Datenzentren aus 6 Ländern zertifiziert werden, 5 weitere Zentren befinden sich im Zertifizierungsprozess.

Eine erste systematische Evaluierung und Auswahl verfügbarer Softwarelösungen für den Einsatz in klinischen Studien in KKS und ZKS wurde mit Unterstützung des KKS-Netzwerks und der TMF bereits vor über 10 Jahren durchgeführt. Auf der Basis

---

<sup>6</sup> Siehe <http://ecrin.org/tools/data-centre-certification>.

dieses Evaluierungsprojektes und mit BMBF-Mitteln konnten zwei Softwarelösungen in KKS/ZKS erfolgreich implementiert werden (MACRO, eResearchNetwork).

In den letzten Jahren haben sich aber sowohl die Anforderungen als auch die angebotenen Softwaresysteme deutlich weiterentwickelt. Vor diesem Hintergrund hat eine Arbeitsgruppe des KKS-Netzwerks unter Federführung des ZKS Köln einen systematischen Anforderungskatalog für CDMS erarbeitet. Mit insgesamt 331 Anforderungen deckt dieser Katalog sowohl web-basierte als auch papier-basierte Studien ab. Neben den Kernanforderungen (eCRF-Implementierung, Anwender- und Rollenverwaltung, Datenerfassung, -prüfung und -korrektur, Berichte, Datenexport, Audit Trail) wurden auch Anforderungen für Zusatzfunktionen (Codierung von Medikamenten und unerwünschten Ereignissen, Berichtsgenerator, Planung und Vorbereitung von Monitoring-Visiten, Randomisation, Medikationsmanagement, SAE-Management, Studienmanagement) formuliert.

Ebenfalls unter Federführung des ZKS Köln wurde auf der Basis der entwickelten CDMS-Requirements eine Softwareevaluierung durchgeführt. Hierzu wurde aus dem Anforderungskatalog ein Fragebogen für eine Herstellerumfrage abgeleitet. Von gut 100 angefragten Herstellern beteiligten sich 21 an einer Umfrage, wobei 19 Rückmeldungen auswertbar waren. Anhand der Antworten wurden 6 potentiell für den Einsatz im Kontext des KKS-Netzwerks geeignete Systeme identifiziert. Diese wurden anschließend von den jeweiligen Herstellern anhand einer detaillierten Agenda im Rahmen einer Web-Präsentation vorgestellt.

Weitere wichtige Rahmenbedingungen für das Datenmanagement in klinischen Studien ergeben sich aus Leitlinien für die Verarbeitung von elektronischen Source-Dokumenten (eSource) [7-9], aus Empfehlungen für die Validierung von CDMS-Systemen (z. B. die Systemvalidierungsunterlagen der TMF<sup>7</sup>) sowie aus Standardisierungsbemühungen in diesem Feld (z. B. CDISC<sup>8</sup>). Unter der Organization for the Advancement of Structured Information Standards (OASIS), einer internationalen, nicht-gewinnorientierten Organisation, die sich mit der Weiterentwicklung von E-Business- und Webservice-Standards beschäftigt, wurde in den letzten Jahren eine Spezifikation für einen Electronic Trial Master File (eTMF) entwickelt und 2016 verabschiedet und veröffentlicht [10]. Ziel dieses offenen und technologieutralen Standards ist die Interoperabilität und der Datenaustausch von Trial Master Files bei klinischen Studien. Nach wie vor ein Problem vieler Standards für klinische Studien ist allerdings ihre mangelnde Akzeptanz und Anwendung.

---

7 Siehe [www.tmf-ev.de/Produkte/P019011](http://www.tmf-ev.de/Produkte/P019011), [www.tmf-ev.de/Produkte/P019012](http://www.tmf-ev.de/Produkte/P019012) und [www.tmf-ev.de/Produkte/P019013](http://www.tmf-ev.de/Produkte/P019013).

8 Siehe [www.cdisc.org](http://www.cdisc.org).

Der immer weiter fortschreitende Einsatz computergestützter Systeme für die Erfassung und Verwaltung von Daten in klinischen Studien stellt die Anwender vor einen nicht zu unterschätzenden Aufwand bezüglich des Qualitätsmanagements und der damit verbundenen notwendigen Validierung der eingesetzten Computersysteme. Immer häufiger werden auch Daten direkt aus den klinischen Systemen der Versorgung übernommen. Die dabei verwendeten Schnittstellen bedürfen dann ebenso einer Validierung. Noch schwieriger wird es, wenn ggf. auch die datenliefernden Systeme aus dem klinischen Umfeld selbst zu validieren sind. Diese unterliegen in der Regel nicht der Verantwortung des jeweiligen Studienzentrums. Welche Vertragskonstellation vorliegt, inwieweit die Regeln einer Datenverarbeitung im Auftrag greifen können und ob eine vertragliche Absicherung der Einhaltung aller GCP-Anforderungen möglich ist, muss dann im Einzelfall geklärt werden.

### Kategorisierung der IT-Landschaft

Eine Abgrenzung der in klinischen Studien eingesetzten IT-Komponenten ist nur schwer möglich. In der Regel überschneiden sich die Funktions- und damit auch Aufgabengebiete der verschiedenen Komponenten. Sehr oft werden aber auch Funktionen durch externe Systemkomponenten übernommen (z. B. Dokumentenmanagementsysteme, CRM-Systeme etc.). Die folgende Darstellung (Abb. 3.1) des aktuellen Bestands fokussiert daher vorrangig Systeme, welche die Kernaufgaben abdecken.

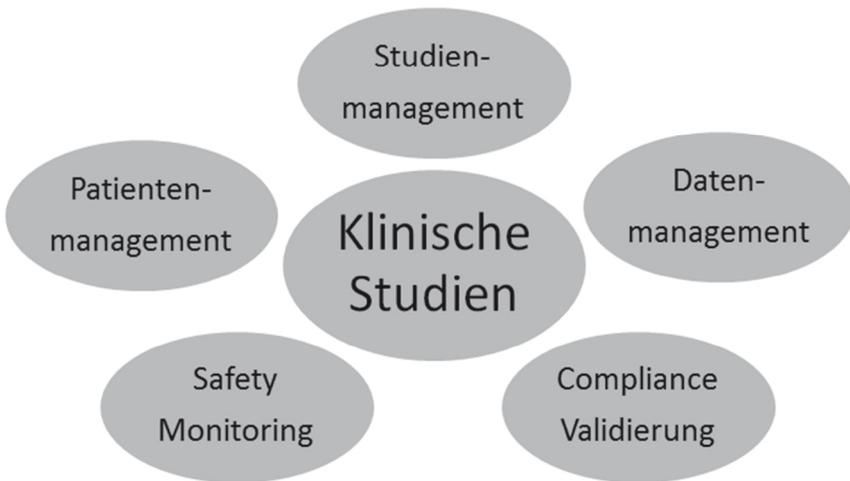


Abbildung 3.1: Funktionsbereiche für die Organisation und Durchführung klinischer Studien.

Generell lassen sich bei den eingesetzten Softwarelösungen kommerzielle Lösungen (Closed-Source), Open-Source-Lösungen sowie Eigenentwicklungen unterscheiden.

- ▶ **Kommerzielle Systeme (Closed Source):** In der Regel fallen für die Lizenzierung derartiger Systeme Kosten an. Dabei unterscheiden sich die Modelle je nachdem ob die Kosten **einmalig** für die Investition, **variabel** je nach dem Umfang der Nutzung (z. B. je Anzahl der Prüfbüros, der Studien, der CRFs etc.) oder **zeitlich** nach der Dauer der Nutzung anfallen. Zusätzlich können auch noch Kosten für Supportverträge anfallen. Vorteile von kommerziellen Lösungen sind stabile, gut unterstützte Lösungen mit einer relativ verlässlichen Langzeitverfügbarkeit sowie definierten Roadmaps. Ein weiterer Vorteil ist der geringere Aufwand im Rahmen der notwendigen Systemvalidierung. Nachteile sind auf der einen Seite die möglichen Kosten, je nach Lizenzmodell, und auf der anderen Seite der fehlende Zugriff auf den Programmcode. So kann es Probleme geben, wenn die Softwareentwicklung z. B. nach einem Firmenverkauf oder einer Insolvenz eingestellt wird.
- ▶ **Open-Source-Lösungen:** Auch bei der Nutzung von Open-Source-Lösungen können je nach Lizenzmodell Kosten für die Nutzung anfallen. Allerdings orientieren sich diese Modelle eher an **variablen** Kosten oder Kosten für Supportverträge. Nicht für alle Open-Source-Lösungen ist ein Supportangebot verfügbar. Oft beschränkt sich der Support auf die Unterstützung aus der Community. Dies kann im Bereich der klinischen Studien ein Problem darstellen. Durch die Verfügbarkeit des Programmcodes kann aber je nach Kompetenz und vorhandenen Ressourcen ein eigener Support organisiert werden. Die Verfügbarkeit des Programmcodes ist somit ein wesentlicher Vorteil beim Einsatz von Open-Source Lösungen. Nachteil ist hier, dass Softwareentwicklungen in diesem Bereich oft durch das Fehlen einer entsprechenden Firmenstruktur von einzelnen Entwicklern abhängen.
- ▶ **Eigenentwicklungen:** Immer noch kommt es im Bereich der klinischen Studien zu Eigenentwicklungen. Der Vorteil dieser Lösungen ist, dass sie passgenau auf die Prozesse im Studienzentrum angepasst werden können. Zudem liegt der Quellcode der Anwendung vor und die notwendige Kompetenz zur Pflege und Weiterentwicklung ist zunächst vorhanden. Nachteil ist aber der sehr hohe Aufwand für Entwicklung und Maintenance, wobei es in der Regel nur einen Anwender der Softwarelösung gibt. Auch ist der Umfang der Systemvalidierung bei Eigenentwicklungen sehr hoch.

Eine weitere Kategorisierung ergibt sich aus dem Funktionsumfang der Softwarelösung. Hier lassen sich auf der einen Seite die Komplettlösungen mit vollständigem Funktionsumfang in einem Softwarepaket und auf der anderen Seite die modularen Lösungen mit einer Verteilung der Funktionen auf verschiedene Softwaremodule unterscheiden.

## Aktueller Bestand

### Datenmanagement

Für die Erfassung und Verwaltung der Patientendaten in klinischen Studien werden Clinical Data Management Systems (CDMS) genutzt.

Eine Umfrage zu den eingesetzten CDMS in 74 Zentren des ECRIN-Netzwerkes im Jahre 2010 zeigte folgende Verteilung in den Kategorien [11]:

- ▶ Kommerzielle Systeme/Closed Source: 59%
- ▶ Eigenentwicklungen/Proprietäre Software: 32%
- ▶ Open Source: 6%
- ▶ Andere: 3%

Eine Mehrzahl der Studienzentren setzte also zum damaligen Zeitpunkt kommerzielle Systeme zur Erfassung und Verarbeitung von Daten in klinischen Studien ein. Während ein Drittel der Zentren Eigenentwicklungen bzw. proprietäre Systeme nutzte, spielen Open-Source-Lösungen nur eine sehr geringe Rolle. Am häufigsten eingesetzt wurden folgende Systeme: MACRO, Capture Systems, eResearchNetwork, CleanWeb, GCP Base, SAS [11].

Eine im Rahmen des EU-geförderten Projektes ECRIN-IA 2015 durchgeführte Umfrage bei Koordinierungszentren für klinische Studien ergab den Einsatz von 24 CDMS-Systemen in 15 Zentren. Mehrheitlich werden kommerzielle Systeme eingesetzt (insgesamt 6 Systeme in 14 Zentren, wobei kein System in mehr als 4 Zentren zur Anwendung kommt). Eine Zunahme ist bei Open-Source-Lösungen zu beobachten (2 Systeme in 4 Zentren: OpenClinica und RedCap), während Eigenentwicklungen deutlich auf dem Rückzug sind (nur noch 4 Lösungen). In den KKS/ZKS eingesetzte kommerzielle CDMS sind MACRO, eClinical, secuTrial, Clincase, Marvin und Trialmaster. Betrachtet man alle sich an der Umfrage beteiligten ECRIN-Zentren (n=53) aus 14 Ländern, so wurden insgesamt 84 Systeme eingesetzt. Hier ergab sich ein Anteil von 48% kommerziellen Systemen [12]. Im Hinblick auf Eigenentwicklungen konnte eine Abnahme (25%) und bei Open-Source Lösungen eine Zunahme (20%) im Vergleich zu der Umfrage von 2010 festgestellt werden. Nur ein System wurde in mehr als 10 Zentren eingesetzt (Macro: 12), lediglich 3 Systeme fanden in mindestens 5 Studienzentren Verwendung (OpenClinica: 7, RedCap: 6, SecuTrial: 5). Bei 41% der Systeme wurde die Installation des CDMS direkt im Studienzentrum, bei 33% im Rechenzentrum der jeweiligen Universitätsklinik oder der Universität vorgenommen. Immerhin

26% der Studienzentren verwendeten extern bei kommerziellen Providern gehostete Systeme. Von allen befragten Zentren bekundeten 62% Interesse an der Evaluierung eines neuen CDMS, was auf Probleme mit der Zufriedenheit der jeweils eingesetzten Systeme hindeutet.

Im Rahmen des bereits beschriebenen CDMS-Projekts des KKS-Netzwerks wurden Softwaresysteme von 19 Providern eruiert, darunter Medidata Rave, TrialMaster, secuTrial, MACRO, OpenClinica und Marvin.

Ein interessantes Modell könnte sich aus einer Entscheidung des National Cancer Instituts (NCI) ergeben: Das NCI hat von der Firma Medidata Solutions Lizenzierungsrechte erworben, die Software Medidata Rave für die Durchführung aller NCI-geförderten klinischen Studien einzusetzen [13]. Ähnliche Bemühungen, wenn auch in kleinerem Rahmen, gibt es z. B. in der Swiss Clinical Trial Organization mit der Einführung von secuTrial [14].

### **Patientenmanagement**

Bei einer Bestandsaufnahme von Softwarelösungen für die Verwaltung von Studienteilnehmern und Untersuchungsterminen im Rahmen eines TMF-Projektes zeigten sich unterschiedliche Anforderungsprofile an derartige Systeme sowie das Fehlen einer einheitlichen Nomenklatur. Im Ergebnis dieser Bestandsaufnahme wurde versucht, eine Terminologie für derartige Verwaltungssysteme zu erstellen, welche eine Einordnung der Systeme erlaubt und in einem nächsten Schritt die Modularisierung und Entwicklung von Schnittstellen ermöglicht.

Insbesondere ist festzustellen, dass sich Lösungen für das sogenannte Customer-Relationship-Management (CRM) oft nicht eignen, insbesondere da notwendige zusätzliche Funktionen zur Unterstützung klinischer Studien fehlen [15]. Auf der anderen Seite gibt es auch Probleme, vorhandene Patientenmanagementsysteme der beteiligten Prüfzentren zu nutzen. Neben den Funktionen für die Umsetzung datenschutzrechtlicher Anforderungen fehlen gerade hier einheitliche Schnittstellen.

### **Studienmanagement**

Ein Clinical Trial Management System (CTMS) erlaubt die Eingabe und Verwaltung der Daten über eine bzw. mehrere Studien. Dies schließt alle relevanten Daten zum aktuellen Status der Studien und ggf. auch der beteiligten Studienzentren (Sites) mit ein. Mit Hilfe eines CTMS wird der Studienmanager und das Studienteam bei der

Planung, der Vorbereitung, der Durchführung und der Auswertung von klinischen Studien unterstützt. Teilweise wird die Datenbank des CTMS mit einem Business-Intelligence-System (SAP, KIS) verknüpft, so dass Studiendaten mit Daten aus anderen Datensystemen im Unternehmen verknüpft und ausgewertet werden können. Der genaue Funktionsumfang eines CTMS differiert von Produkt zu Produkt. Zudem sind die Anforderungen an jede Implementierung unterschiedlich, so dass viele CTMS-Lösungen über tiefgreifende Konfigurationsmöglichkeiten verfügen. Der Funktionsumfang eines CTMS überschneidet sich mitunter mit anderen in der klinischen Forschung eingesetzten Softwaresystemen wie z. B. dem Studiendatenmanagement, dem Safetymanagement bzw. Monitoring oder auch dem Patientenmanagement.

Im Rahmen einer Untersuchung vorhandener Lösungen am ZKS Leipzig zeigte sich eine sehr heterogene Landschaft der eingesetzten Systeme. Bei den Koordinierungszentren für Klinische Studien werden überwiegend Eigenentwicklungen eingesetzt. Diese unterscheiden sich sehr stark in ihrem Funktionsumfang. Außerhalb des Umfeldes der KKS gibt es im nichtkommerziellen Bereich an einigen Universitätskliniken ebenfalls Eigenentwicklungen. Hier sind besonders das in Ulm eingesetzte Trial Management System und das am Clinical Trial Center Aachen eingesetzte Study Management Tool<sup>9</sup> hervorzuheben.

Im Bereich der Open-Source-Lösungen gibt es neben der caBIG Clinical Trial Suite praktisch keine fertigen Lösungen. Die caBIG Clinical Trial Suite (CTS) wird allerdings schon seit mehreren Jahren nicht mehr weiterentwickelt und es gibt keine Supportleistungen. Zudem ist das CTS sehr eng in das caBIG-Framework eingebunden, so dass ein Einsatz des CTS ohne die Nutzung weiterer caBIG-Produkte nicht zielführend ist.

Im kommerziellen Bereich gibt es im Gegensatz dazu eine Vielzahl an Lösungen. Hervorzuheben sind hier beispielsweise:

- ▶ Allegro CTMS (<http://forterresearch.com/ctms-allegro/>)
- ▶ Bio-Optronics: Clinical Conductor (<http://bio-optronics.com/clinical-conductor/>)
- ▶ DDI: mPORTAL (<http://www.ddismart.com/portal-clinical-collaboration>)
- ▶ ACSIS (<http://acsis201.de/acsis/studienverwaltung/>)

Diese Systeme werden aber in der nichtkommerziellen klinischen Studienforschung aufgrund der erheblichen Lizenzkosten kaum eingesetzt.

---

<sup>9</sup> Siehe Info-Kasten zu dieser Lösung.

## Transparenz bei klinischen Studien

### Metadaten von Studienobjekten

Eine zunehmende Anzahl von Interessensgruppen (z. B. Forscher, Patienten, Förderer, Ministerien) fordert eine stärkere Transparenz in der klinischen Forschung. In einer Serie von Artikeln hat sich 2014 die Zeitschrift Lancet diesem Thema unter der Überschrift „increasing value, reducing waste“ angenommen [vgl. 16]. In diesem Kontext wurden viele Initiativen ins Leben gerufen und zahlreiche unterschiedliche Repositories entwickelt, die Daten und Dokumente klinischer Studien zur Verfügung stellen. Diese sogenannten Datenobjekte klinischer Studien können z. B. das Studienprotokoll, den Statistischen Analyseplan, den Datenmanagementplan, die Case Report Forms oder auch die Datenbank mit individuellen Daten von Studienteilnehmern umfassen. Um eine bestmögliche Nutzung dieser Ressourcen sicherzustellen, ist ein weit gehender Konsens über ein einfaches und konsistentes Metadaten-Schema notwendig.

Das im Rahmen des EU-Programms Horizon 2020 geförderte CORBEL-Projekte<sup>10</sup> hat in WP3 ein solches Metadaten-Schema entwickelt, das auf dem weit verbreiteten DataCite Standard basiert und die spezifischen Belange der klinischen Forschung berücksichtigt [17]. Das Schema unterstützt die Identifikation von klinischen Studien mit einem Link zu Studienregistern, charakterisiert Datenobjekte, die zu einer klinischen Studie gehören und beschreibt die Lokation und den Zugang zu diesen Datenobjekten. Das Schema stellt eine natürliche Erweiterung eines weit verbreiteten Standards dar und schließt eine Lücke, die von anderen Standards (z. B. BRIDG, CDISC) nicht abgedeckt wurde.

### Metadata Repository

Ein Metadata Repository, welches eine klinische Studie mit Ihren verschiedenen Datenobjekten zusammenführt, stellt eine wichtige Infrastruktur zur Zusammenführung und sekundären Nutzung der gesamten zu einer klinischen Studie gehörigen Information dar. Geeignete Tools und Portale sollten in einem ersten Schritt die Identifikation einer klinischen Studie anhand der Metadaten und mit Unterstützung durch Studienregister ermöglichen. In einem zweiten Schritt sollte dann die Verlinkung zu den Metadaten der zu dieser Studie gehörenden Datenobjekte erfolgen. Ein Konzept für ein solches Metadata Repository wurde ebenfalls im Rahmen des CORBEL-Projektes erarbeitet [18]. Eine prototypische Umsetzung eines Metadaten Repositories stellt OpenTrials<sup>11</sup>

---

<sup>10</sup> Siehe <http://www.corbel-project.eu/home.html>.

<sup>11</sup> Siehe <http://opentrials.net>.

dar, welches in einer öffentlichen Vorversion vorliegt [19]. Die Metadaten zu Studien- und anderen medizinischen Formularen finden sich in dem Portal „Medical Data Models“ der Universität Münster [20].<sup>12</sup>

### **Zugang zu individuellen Daten von Studienteilnehmern**

Mittlerweile gibt es eine breite Zustimmung für die Öffnung des Zugangs zu individuellen Daten von Studienteilnehmern für sekundäre Analysen. Zeitschriften und Förderer fordern zunehmend einen Data Sharing Plan, indem dieser Prozess dargestellt werden soll. Die pharmazeutische Industrie hat sich ebenfalls diesem Thema zugewendet und unterstützt Repositories, die einen Zugang zu Studien der pharmazeutischen Industrie ermöglichen (z. B. Yale University Open Data Access, YODA; ClinicalStudyDataRequest.com). Im Jahre 2016 wurde ein Statement des International Committee of Medical Journal Editors (ICMJE) publiziert, in dem die Autoren aufgefordert werden, anonymisierte individuelle Daten von Studienteilnehmern mit anderen Forschern zu teilen [21]. In diesem Zusammenhang wurde eine Vielzahl von Repositories entwickelt, die die Nutzung dieser Daten für sekundäre Analysen unterstützen sollen. Einige dieser Repositories sind generisch (z. B. DRYAD, Zenodo, Dataverse), andere sind krankheitsspezifisch (z. B. NIDA Data Share, Repository of Registered Migraine Trials – RReMIT). Ein generisches Repository für Daten kommerzieller und nicht-kommerzieller klinischer Studien wird derzeit vom Multi Regional Clinical Trials Center (MRCT) an der Harvard University entwickelt (Vivli) [22]. Das bereits erwähnte Projekt CORBEL beschäftigt sich damit, ein Konsens-Dokument für den Zugang zu individuellen Daten nicht-kommerzieller klinischer Studien zu erarbeiten. Dieses Dokument wird in der ersten Hälfte 2017 vorliegen und auch einen Überblick zu vorhanden Repositories enthalten.

---

<sup>12</sup> Siehe auch das Kapitel zu Metadaten und Metadata Repositories in diesem Bericht.

# Study Management Tool des Clinical Trial Centers Aachen

## CTMS-Grundfunktionen

Das Study Management Tool (SMT) des Clinical Trial Centers Aachen (CTC-A) ist eine Web-basierte Datenbank zum Management klinischer Studien, die zusammen mit dem Institut für Medizinische Informatik der RWTH Aachen entwickelt wurde.

Die SMT-Kerndatenkategorien sind (in Klammern die Daten für Aachen, Stand 20.11.2015):

- ▶ Study: alle relevanten Informationen und Referenznummern, sowie Klassifizierungen (1466, davon 337 rekrutierend)
- ▶ Person: alle Personen mit ihren studienspezifischen Rollen (961, davon 217 mit Login)
- ▶ Department: alle Institutionen (568, davon 65 RWTH Aachen), wobei eine Person genau einem Department zugeordnet ist und hierüber das mehrschichtige Zugriffsmodell modelliert ist (no, own, department, all, master)
- ▶ Patient: alle Patienten in den Studien (11.598, davon 2.526 mit TMF-PID integriert)
- ▶ Efforts: Task-spezifische Aufwandserfassung (124 Personen, 126.151 Einträge)
- ▶ Documents: Studienspezifisches Dokumenten- und Versionsmanagement und SOP-System als Basis der CTC-A ISO 9001 Zertifizierung (2.042, davon 1.869 aktuell)
- ▶ Costs: Vollkostenkalkulationstool mit allen Tabellen aus GoÄ, TVL, etc.
- ▶ Invoices: Automatische Generierung SAP-konformer Rechnungen, studienspezifischer Materialkataloge und interner Verteilungsschlüssel und Übertragung an die Krankenhaus-IT

## CTMS-Systemintegration

Aufgrund der seit 2010 kontinuierlich erfolgten Weiterentwicklung des SMT können heute zahlreiche Web Services und Schnittstellen angeboten werden, mit

deren Hilfe das SMT die Rolle eines Kommunikationsservers zwischen derzeit folgenden Komponenten einnimmt (vgl. Abb. 3.2):

- ▶ TMF-PID-Generator zur Pseudonymisierung
- ▶ R als statistisches Programm zur Randomisierung mit externer Einsprungstelle für multizentrische Studien
- ▶ OpenClinica als Electronic Data Capture System (EDC) für Studiendaten [23]
- ▶ Rare Disease Registry (RDR) Framework für medizinische Register [24]
- ▶ DICOM4CHEE als Picture Archiving and Communication System (PACS) für Binärdaten [25]
- ▶ ImageJ/Matlab Application Programming Interface (API) für Bild- und Signalverarbeitung für die Bestimmung von Surrogate Endpoints (Bild- und Signalbasierte Biomarker) in klinischen Studien [26]
- ▶ iOS / Android Apps zur Anzeige der tagesaktuellen Patienteneinschlussstatistiken

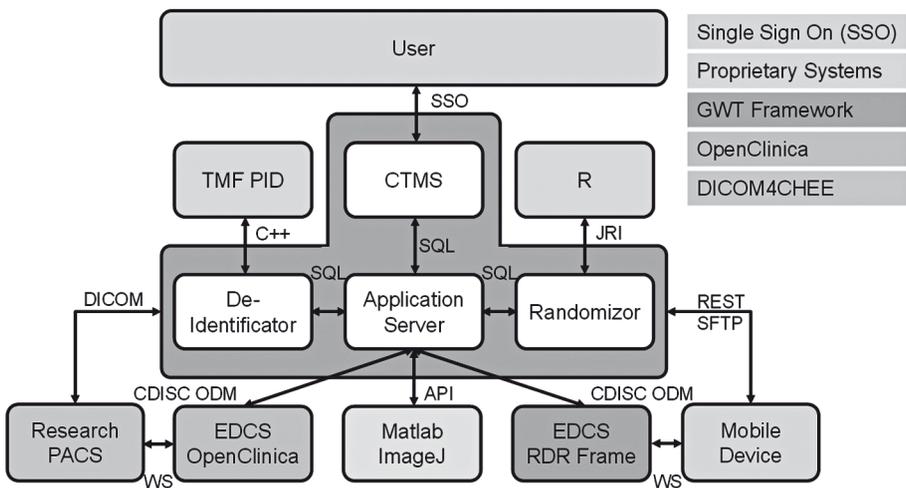


Abbildung 3.2: Systemintegration und Schnittstellen des Study Management Tool des Clinical Trial Centers Aachen als beispielhaftes Clinical Trial Management System (CTMS). Beschreibung siehe Infobox oben.

## Compliance und Validierung computergestützter Systeme in klinischen Studien

Beim Einsatz von IT-Systemen in der klinischen Forschung ist eine wesentliche Grundlage für die Sicherstellung der Compliance die Validierung derartiger Systeme. Für die Unterstützung bei der Durchführung der Validierung existiert als eine erste Grundla-

ge das Validierungspaket der TMF.<sup>13</sup> Dieses enthält eine umfassende Sammlung an Musterdokumenten, Checklisten und Vorlagen. Zudem gibt es ein Schulungskonzept, welches auf modularer Basis Schulungen für verschiedene Zielgruppen ermöglicht. Die Erfahrungen in den letzten Jahren haben aber gezeigt, dass der Umfang der vorliegenden Dokumente einerseits zu groß ist, um sie praktikabel einsetzen zu können. Andererseits sind die Dokumente oft inkonsistent (unterschiedliche Sprachen, Bezeichnungen, Dopplungen). In einem Projekt der TMF findet aktuell eine Überarbeitung der Dokumente des Validierungspaketes statt.

Ein weiteres Problem ist ein fehlendes Konzept zur Selbstkontrolle und Überprüfung des validen Zustandes. Hilfreich wäre ein Vorgehen, welches die Nutzer im Verlauf der Validierung früh auf eventuelle Fehler in der Durchführung und Lücken in der Dokumentation hinweist. In der Regel sind Findings, die erst im Rahmen von behördlichen Kontrollen auftauchen, dann oft nur mit einem erhöhten Aufwand zu beheben.

Aber auch auf der Seite der Behörden bestehen einige Unsicherheiten bezüglich der zu erwartenden Voraussetzungen bei den zu prüfenden universitären Einrichtungen. Ein entsprechend abgestimmtes Verfahren zur Selbstkontrolle bzw. zur Überprüfung des validen Zustandes – unabhängig von behördlichen Inspektionen – würde hier auf beiden Seiten für mehr Sicherheit sorgen.

## Bewertung und Handlungsbedarf

Es besteht nach wie vor großer Bedarf an bezahlbarer professioneller Software im Non-Profit-Bereich. Wissenschaftsgesteuerte klinische Studien (IITs) sind ein unverzichtbares Element der klinischen Forschung. Mit ihren methodischen Anforderungen stellen sie jedoch nach wie vor eine große Herausforderung für akademische Studienzentren dar. Aufgrund der Finanzlage sind größere Investitionen in Studiensoftware an den meisten Standorten nicht oder nur bedingt möglich. Aus diesem Grunde wird derzeit auch an Alternativen gearbeitet. Im Rahmen des von der EU geförderten Projekts ECRIN-IA hat die European Organisation for Research and Treatment of Cancer (EORTC) mit Beteiligung einzelner KKS mit der Entwicklung eines CDMS für den akademischen Bereich (VISTA) begonnen. VISTA ist eine Weiterentwicklung einer bestehenden und routinemäßig eingesetzten Software und wird nach Fertigstellung die relevanten CDISC-Standards unterstützen sowie die Anforderungen nach GCP und 21 CFR Part 11 erfüllen. Aus technischer Sicht ist VISTA ein beeindruckendes Tool mit einer sehr großen Funktionalität. Allerdings konnte die Softwareentwicklung während

---

13 Siehe [www.tmf-ev.de/Produkte/P019011](http://www.tmf-ev.de/Produkte/P019011), [www.tmf-ev.de/Produkte/P019012](http://www.tmf-ev.de/Produkte/P019012) und [www.tmf-ev.de/Produkte/P019013](http://www.tmf-ev.de/Produkte/P019013).

der Projektdauer nicht abgeschlossen werden. Entsprechend wurde die Testung und Validierung des Tools durch die unabhängigen VISTA-Tester nur mit einer nicht finalen Version vorgenommen. Die Pilotversion von VISTA Trials 1.0 wurde am 11.5.2016 nach umfassender Testung fertiggestellt. Mit der Testung von VISTA Trials in einer realen Studie (EORTC-Studie 1527 Dream) wurde in Kooperation mit der Japan Clinical Oncology Group begonnen. Es ist projektiert, dass die komplette Validierung und Dokumentation bis Dezember 2016 abgeschlossen ist, so dass die validierte Version von VISTA Trials 1.0 im Januar 2017 ausgeliefert werden kann.

Auf der Basis der Ergebnisse des TMF-Projekts zum Probanden- und Einwilligungsmanagement sollte unbedingt eine einheitliche Terminologie derartiger Systeme und Funktionen, abgestimmt mit internationalen Ergebnissen auf diesem Gebiet, entwickelt werden. Darauf aufbauend müssen einheitliche Schnittstellen zu anderen Systemen definiert werden, die idealerweise internationalen Standardvorgaben für Schnittstellen im Gesundheits- oder Forschungsbereich entsprechen.

Im Bereich der Systeme für das Management klinischer Studien ist die Lage momentan so, dass es kaum vorhandene Lösungen im Open-Source-Bereich gibt und die einzelnen Produkte zudem noch einen unzureichenden Funktionsumfang bieten. Die an den Studienzentren eingesetzten Eigenentwicklungen sind oft entweder stark in die vorhandene Infrastruktur vor Ort integriert oder bilden nur einen Teil der notwendigen Funktionen ab. Lediglich die Lösungen aus Ulm (Ulmer Trial Management System) und Aachen (Study Management Tool) scheinen hier eine interessante Option darzustellen. Beide Lösungen bedürfen aber neben einer entsprechenden Lizenzgestaltung auch eines Konzepts für den notwendigen Support und die langfristige Pflege und Weiterentwicklung.

Die immer komplexer werdenden Infrastrukturen mit ihrer verteilten Datenhaltung erfordern einen immer höheren Aufwand bei der Herstellung und Aufrechterhaltung eines validen Zustands. Das Qualitätsmanagement in diesem Bereich stellt gerade akademische Einrichtungen vor große Herausforderungen. Diese bestehen darin, a.) qualifiziertes Personal auszubilden, zu akquirieren und trotz z. T. unterdotierter Zeitverträge zu halten, b.) verteilte IT-Infrastrukturen mit diversen Partnern innerhalb und außerhalb der eigenen Institution zu managen und dabei c.) zunehmend verschärfte regulatorische Vorgaben einzuhalten. Ein Austausch der akademischen Zentren zu diesen Themen wird unbedingt benötigt.

Im Bereich der Validierung der IT-Systeme wäre ebenfalls eine bessere Kommunikation und Abstimmung mit den zuständigen Behörden hilfreich und würde den Validierungsaufwand in den Forschungsverbänden verringern. Begleitend hierzu muss es geeignete

Beratungs- und Auditkonzepte geben. Grundlage hierfür könnte das Validierungspaket der TMF sein, welches sich momentan in Überarbeitung befindet. Insbesondere ist es notwendig, entsprechende Verfahren für eine Selbstkontrolle und Überprüfung des jeweiligen validen Zustandes zu entwickeln und dieses mit den Behörden, soweit möglich, abzustimmen. Ein mögliches Ziel wäre ein mit allen relevanten Behörden abgestimmtes Validierungspaket, ähnlich der mit den zuständigen Aufsichtsbehörden verhandelten und von diesen empfohlenen generischen Datenschutzkonzepte der TMF [27].

Die im Bereich der modernen Softwareentwicklung immer häufiger eingesetzten unterschiedlichen Methoden der Programmierung (z. B. SCRUM) erfordern zudem praktikable Validierungsverfahren. Auch der Umfang der Validierung und die Zuständigkeiten bei der automatischen Datenübernahme aus klinischen Informationssystemen sind grundlegend zu klären. Zudem gibt es beim Einsatz cloud-basierter Dienste noch kaum Erfahrungen mit der notwendigen Validierung der Systeme. Hier ist in Zusammenarbeit mit den Behörden eine entsprechende Vorgehensweise zu entwickeln.

Das große Forschungsinteresse an den hoch qualitätsgesicherten Daten aus klinischen Studien führt zunehmend zu der Anforderung, diese auch nach Abschluss einer Studie z. B. im Rahmen übergreifender Register oder Forschungsdatenbanken weiter zu verwerten. Von besonderer Bedeutung ist in diesem Zusammenhang die Transparenzinitiative der Europäischen Zulassungsbehörde EMA, die einen öffentlichen Zugang zu der EU-Datenbank mit Daten klinischer Studien mit Arzneimitteln gewährleisten will (EU-Verordnung Nr. 536/2014; [28]). Ein Interesse daran ist sowohl in der Industrie als auch in der akademischen Forschung vorhanden. Dieses Interesse kann sich z. B. auch zusätzlich auf die im Rahmen klinischer Studien erhobenen und ggf. weiter verwertbaren Bioproben beziehen. Nicht zuletzt diese Anforderungen haben auch zu den neuen modularen Datenschutzkonzepten der TMF geführt, die hierfür die notwendigen rechtlichen, technischen und organisatorischen Rahmenbedingungen beschreiben [27]. Eine Zusammenführung von Daten aus unterschiedlichen Studien ist jedoch nur dann sinnvoll möglich, wenn diese standardisiert definiert und dokumentiert werden. Insgesamt ist daher ein zunehmender Bedarf an Standardisierung der Datenerhebung zu konstatieren, der sich insbesondere auch auf standardisierte und automatisiert auswertbare Metadaten erstreckt.

Mittelfristig gilt es somit die folgenden Bereiche zu stärken:

- ▶ Bessere Unterstützung von Standards durch CDMS (z. B. CDISC, OASIS) vor allen Dingen vor dem Hintergrund der Empfehlungen der FDA und japanischer Zulassungsbehörden.

- ▶ Interoperabilität von CDMS mit EHR (abgeschlossene EU-Projekte: Transform, EHR4CR [29], p-medicine; abgeschlossenes BMBF-Projekt KISREK [30])
- ▶ Schnittstellen von CDMS zu Biobanken, Bilddaten, genetischen Daten, etc. (abgeschlossene EU-Projekte: BioMedBridges, p-medicine). In diesem Bereich sind dringend nationale Koordinierungsaktivitäten notwendig. Hierfür sollte die bestehende Kooperation des KKS-Netzwerks mit der TMF verstärkt genutzt werden.
- ▶ Die Vorhaben zur Bereitstellung standardisierter und qualitätsgesicherter Datenelemente mit maschinenlesbaren und automatisiert verwertbaren Metadaten müssen fortgeführt und nach Möglichkeit verzahnt werden (z. B. CDISC-SHARE, BMBF-Projekt zum Aufbau eines MDR<sup>14</sup> und „Medical Data Models“ der Universität Münster<sup>15</sup>). CDMS müssen Schnittstellen zum Im- und Export von Datenelementen samt standardisierter Metadaten anbieten. Hierbei sind vor allen Dingen internationale Entwicklungen zu berücksichtigen (z. B. Metadatenschema auf der Basis von DataCite, OpenTrials).
- ▶ Entwicklung und Abstimmung von Anforderungen und Leitlinien zur Nutzung cloud-basierter Services für klinische Studien (z. B. CDMS als SaaS). Hier sollte auf den Ergebnissen des Cloud Workshops von ECRIN [1] und den Diskussionen mit der Zulassungsbehörde EMA aufgesetzt werden. Von Bedeutung könnte auch ein neues EU-gefördertes Projekt werden, welches nächstes Jahr starten wird (The European Open Science Cloud for Research Pilot Project).<sup>16</sup>
- ▶ Definition einer einheitlichen Terminologie zu Probanden – und Einwilligungsverwaltungssystemen mit einer Abgrenzung des Funktionsumfangs und der Anforderungen. Auf der Basis dieser Terminologie und Spezifikationen müssen offene Schnittstellen zur Integration derartiger Systeme in die vorhandene Infrastruktur der klinischen Studien entwickelt werden.

## Literatur

1. Ohmann, C., Canham, S., Danielyan, E., Robertshaw, S., Legré, Y., Clivio, L., Demotes, J., *Cloud computing and clinical trials: report from an ECRIN workshop*. *Trials*, 2015. **16**: S. 318-321.
2. PhUSE *Cloud Services. A Framework for Adoption in the Regulated Life Sciences Industry*. 2015. Pharmaceutical User Software Exchange - Cloud Adoption Working Group, <https://s3-us-west-2.amazonaws.com/phuse/public/PhUSE+Cloud+Doc+13-Nov-2015.pdf> (Abruf: 2016-11-18).

---

14 [www.tmf-ev.de/mdr](http://www.tmf-ev.de/mdr)

15 <http://medical-data-models.org>

16 Vgl. hierzu auch das Kapitel „Cloud-Computing in der medizinischen Forschung“ im vorliegenden Report.

3. EMA *ICH Topic E 6 (R1) - Guideline for Good Clinical Practice* 2002. European Medicines Agency, [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500002874.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf) (Abruf: 2016-11-18).
4. Ohmann, C., Canham, S., Cornu, C. et al., *Revising the ECRIN standard requirements for information technology and data management in clinical trials*. *Trials*, 2013. **14**: S. 97.
5. Ohmann, C., Kuchinke, W., Canham, S. et al., *Standard requirements for GCP-compliant data management in multinational clinical trials*. *Trials*, 2011. **12**: S. 85 - 93.
6. Canham, S., Clivio, L., Cornu, C. et al. *Requirements for Certification of ECRIN Data Centres, with Explanation and Elaboration of Standards, Version 3.0*. 2015, <http://zenodo.org/record/32690> (Abruf: 2016-11-18).
7. CDISC *Electronic Source Data Interchange (eSDI) Group. Leveraging the CDISC Standards to Facilitate the use of Electronic Source Data within Clinical Trials* 2006. Clinical Data Interchange Standards Consortium, [http://www.cdisc.org/system/files/all/cdisc\\_blog\\_entry/application/pdf/esdi.pdf](http://www.cdisc.org/system/files/all/cdisc_blog_entry/application/pdf/esdi.pdf) (Abruf: 2016-11-18).
8. EMA *Reflection paper on expectations for electronic source documents used in clinical trials*. 2007. European Medicines Agency, [http://www.emea.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/10/WC500004385.pdf](http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500004385.pdf) (Abruf: 2016-11-18).
9. FDA *Guidance for Industry. Electronic Source Data in Clinical Investigations* 2013. U.S. Food and Drug Administration, <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf> (Abruf: 2016-11-18).
10. OASIS *Electronic Trial Master File (eTMF) Specification Version 1.0, Committee Specification 01*. 2016. OASIS, <http://docs.oasis-open.org/etmf/etmf/v1.0/cs01/etmf-v1.0-cs01.pdf> (Abruf: 2016-11-18).
11. Kuchinke, W., Ohmann, C., Yang, Q. et al., *Heterogeneity prevails: the state of clinical trial data management in Europe - results of a survey of ECRIN centres*. *Trials*, 2010. **11**: S. 79 - 89.
12. Ohmann, C., Canham, S., Vogt, A., Ruyskart, P., Demotes-Mainard, J. *Survey on the deployment model of a Clinical Data Management System under development (VISTA Trials)*. 2016. Zenodo, <https://zenodo.org/record/44821> (Abruf: 2016-11-18).
13. NCI *Implementation of the NCI's National Clinical Trials Network. NCI Perspective Article*. 2014, <http://www.cancer.gov/news-events/press-releases/2014/nctnlaunch> (Abruf: 2016-11-18).

14. SCTO *Swiss Clinical Trial Organisation – Data Management. Promoting a nationally harmonised culture of high-quality clinical research.* 2015. Swiss Clinical Trial Organisation, <https://www.scto.ch/de/publications.html> (Abruf: 2016-11-18).
15. Schwanke, J., Rienhoff, O., Schulze, T.G., Nussbeck, S.Y., *Suitability of Customer Relationship Management Systems for the Management of Study Participants in Biomedical Research.* *Methods of Information in Medicine*, 2013. **52**(4 2013): S. 340 - 350.
16. Chan, A.-W., Song, F., Vickers, A. et al., *Increasing value and reducing waste: addressing inaccessible research.* *The Lancet*, 2014. **S0140-6736**(13): S. 62296 - 62295.
17. Canham, S., Ohmann, C., *A metadata schema for data objects in clinical research.* *Trials*, 2016. **17**(1): S. 557.
18. Ohmann, C., Canham, S., Kuchinke, W. et al., *A metadata repository for clinical trials and linked data objects: requirements and issues.* *BMJ*, submitted.
19. Goldacre, B., Gray, J., *OpenTrials: towards a collaborative open database of all available information on all clinical trials.* *Trials*, 2016. **17**: S. 164.
20. Dugas, M., Neuhaus, P., Meidt, A., Doods, J., Storck, M., Bruland, P., Varghese, J., *Portal of medical data models: information infrastructure for medical research and healthcare.* *Database*, 2016. **2016**.
21. Taichman, D.B., Backus, J., Baethge, C. et al., *Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors.* *Ann Intern Med*, 2016.
22. Bierer, B.E., Li, R., Barnes, M., Sim, I., *A Global, Neutral Platform for Sharing Trial Data.* *N Engl J Med*, 2016. **374**(25): S. 2411-2413.
23. Haak, D., Samsel, C., Gehlen, J., Jonas, S., Deserno, T.M., *Simplifying electronic data capture in clinical trials: workflow embedded image and biosignal file integration and analysis via web services.* *J Digit Imaging*, 2014. **27**(5): S. 571-580.
24. Deserno, T.M., Haak, D., Brandenburg, V., Deserno, V., Classen, C., Specht, P., *Integrated image data and medical record management for rare disease registries. A general framework and its instantiation to the German Calciphylaxis Registry.* *J Digit Imaging*, 2014. **27**(6): S. 702-713.
25. Haak, D., Page, C.E., Reinartz, S., Krüger, T., Deserno, T.M., *DICOM for Clinical Research: PACS-Integrated Electronic Data Capture in Multi-Center Trials.* *J Digit Imaging*, 2015. **28**(5): S. 558-566.
26. Sárándi, I., Claßen, D.P., Astvatsatourov, A., Pfaar, O., Klimek, L., Mösges, R., Deserno, T.M., *Quantitative conjunctival provocation test for controlled clinical trials.* *Methods Inf Med*, 2014. **53**(4): S. 238-244.

27. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
28. Banzi, R., Bertele, V., Demotes-Mainard, J., Garattini, S., Gluud, C., Kubiak, C., Ohmann, C., *Fostering EMA's transparency policy*. Eur J Intern Med, 2014. **25**(8): S. 681-684.
29. Doods, J., Bache, R., McGilchrist, M., Daniel, C., Dugas, M., Fritz, F., *Piloting the EHR4CR Feasibility Platform across Europe*. Methods of Information in Medicine, 2014. **53**(4): S. 264-268.
30. Schreiweis, B., Trinczek, B., Köpcke, F. et al., *Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials*. Int J Med Inform, 2014. **83**(11): S. 860-868.



# Register und Kohorten

Jürgen Stausberg, Udo Altmann<sup>1</sup>

Mitarbeit: Karoline Buckow<sup>2</sup>, Stefan Ostrzinski, Wolfgang Hoffmann<sup>3</sup>

## Anwendungsbereich

Register und Kohorten sind in komplexe Vorhaben der patienten- bzw. bevölkerungsorientierten medizinischen Forschung eingebettet [1]. Bei diesen Vorhaben bildet die Erhebung gesundheitsrelevanter Daten interessierender Kollektive die Grundlage für patientenübergreifende Auswertungen. Neben der Verwendung für die medizinische Forschung leisten die Vorhaben auch einen Beitrag zur Gesundheitsversorgung, z. B. im Bereich der Qualitätssicherung und des Qualitätsmanagements [2]. Register und Kohorten stellen daher für viele der in der TMF zusammengeschlossenen Verbände der vernetzten medizinischen Forschung ein zentrales Element dar. Die TMF hat deshalb bereits mehrfach die Register ihrer Mitgliedsverbände analysiert. Einige der hierdurch gewonnenen Ergebnisse von 2012 werden im Rahmen des Kapitels wiedergegeben. Erstmals mit dem IT-Review von 2015 wurden gezielt für Register und Kohorten angebotene Software-Produkte vorgestellt. Diese Übersicht soll in den nächsten Jahren ausgebaut und vervollständigt werden. Mit ihrer gesetzlichen Verankerung und flächendeckenden Bedeutung nehmen die Krebsregister in Deutschland eine Sonderstellung ein und werden in einem eigenen Abschnitt ausführlich dargestellt.

Register sind eine Form der medizinischen Dokumentation. Das Memorandum des Deutschen Netzwerks Versorgungsforschung (DNVF) kennzeichnet sie als eine „möglichst aktive, standardisierte Dokumentation von Beobachtungseinheiten zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen, für die ein präziser Bezug zur Zielpopulation transparent darstellbar ist“ [3, S. 825]. Im Regelfall schließen Register Personen mit einem spezifischen, gesundheitsrelevanten Phänomen ein, vielfach einer Erkrankung, alternativ auch einer Maßnahme. Laut Memorandum des DNVF können Register einer Vielzahl von Zielen dienen, z. B. der Beschreibung epidemiologischer Zusammenhänge und Unterschiede, der Qualitätssicherung und -verbesserung, der Unterstützung klinischer Forschung, der Evaluation und dem Monitoring der Patientensicherheit, der Wirksamkeitsprüfung in der Versorgungsroutine, der ökonomischen Evaluation, der Mindestmengenforschung sowie der Unterstützung der Versorgungsplanung. Die US-amerikanische Agency for Healthcare Research and Quality (AHRQ)

1 Udo Altmann hat den neuen Abschnitt zu Krebsregistern verfasst; Jürgen Stausberg hat die Textteile aus dem IT-Report 2014 aktualisiert und die Erhebung der Software für Register und Kohorten durchgeführt. An der vorangehenden Fassung dieses Kapitels war Claudia Michalik, ZKS Köln, als Autorin beteiligt.

2 Karoline Buckow hat nach Rücksprache mit der DMSG die Infobox zum Deutschen MS-Register eingebracht.

3 Stefan Ostrzinski und Wolfgang Hoffmann haben die Infobox zur Nationalen Kohorte beigeleitet.

nennt in Bezug auf Outcome-orientierte Register folgende Ziele [1]: Beschreibung des natürlichen Krankheitsverlaufs, Feststellung der klinischen und/oder ökonomischen Wirksamkeit, Bewertung von Sicherheit und Schaden, Messung und Verbesserung der Versorgungsqualität, Gesundheitsüberwachung sowie Krankheitsbekämpfung.

Im Unterschied zu Registern erfassen Kohorten einen Ausschnitt der Bevölkerung, in dem sich unterschiedliche Risiken im Hinblick auf interessierende Ereignisse wiederfinden. Das Deutsche Netzwerk Evidenzbasierte Medizin (DNEbM) definiert eine Kohorte als „eine Gruppe von Personen mit gemeinsamen Charakteristika (z. B. Alter, Geschlecht, Gesundheitszustand), die in klinischen und epidemiologischen Studien beobachtet werden“ und eine Kohortenstudie als „eine vergleichende Beobachtungsstudie, in der Personen (Kohorte) mit bzw. ohne eine Intervention/Exposition (zu der sie nicht von dem Studienarzt zugeteilt wurden) über einen definierten Zeitraum beobachtet werden, um Unterschiede im Auftreten der Zielerkrankung festzustellen“ [4]. Das Einschlusskriterium bei Registern entspricht vielfach dem Outcome bei krankheitsbezogenen Kohorten.

Data Repositories sind eine jüngere Entwicklung, die sowohl von einzelnen Einrichtungen als auch von Verbänden vorangetrieben werden und deren Ziel die Bereitstellung von Datensätzen für vorab nicht bekannte Fragestellungen ist. Dazu nehmen Data Repositories Datenbestände von anderen Vorhaben, z. B. Register und Kohorten, auf und stellen aus diesen auf Anfrage Daten zur Verfügung, sei es von Subgruppen, kompletten Vorhaben oder vorhabenübergreifend. Ein Data Repository kann als Dienst, der Daten aus unterschiedlichen Quellen zugänglich macht, definiert werden [5]. Zunehmend verschwimmen die Unterschiede zwischen Registern, Kohorten und Data Repositories. An dieser Stelle sei beispielhaft auf die Angaben zum Phänotyp in Biobanken hingewiesen. Der dort entstehende Datenbestand weist Eigenschaften aller drei der hier behandelten Typen medizinischer Datenbestände auf. Angesichts einer breiter werdenden Diskussion um Data Repositories, werden diese beginnend mit dem IT-Review von 2015 im größeren Kontext des Forschungsdatenmanagements dargestellt.<sup>4</sup>

## Rahmenbedingungen

Für Register und Kohorten existieren nur wenige spezifische gesetzliche Regelungen. Umfassend geregelt ist der Bereich epidemiologischer Krebsregister in Gesetzen der Bundesländer. Durch das Gesetz zur Einführung klinischer Krebsregister ist auch für diesen Bereich eine gesetzliche Grundlage entstanden (s. unten). Weitere Register-

---

<sup>4</sup> Siehe hierzu das Kapitel „Forschungsdatenmanagement“ im vorliegenden Report.

ähnliche Datenbestände werden durch gesetzliche Regelungen zur Todesursachenstatistik, zum Infektionsschutz, zur Pharmakovigilanz und anderen Sachverhalten, die weitgehend der Gesundheitsberichterstattung zugeordnet werden können, definiert. Im Bereich der Krankenversorgung sind vorrangig die Regelungen zur Qualitätssicherung von Bedeutung, sei es die sektorübergreifende Qualitätssicherung nach § 137 Sozialgesetzbuch V (SGB V), die Qualitätssicherung bei Disease-Management-Programmen oder Vereinbarungen der Selbstverwaltungspartner in Bezug auf einzelne Maßnahmen wie z. B. der Vorsorgekoloskopie. Für den Betrieb von Registern und Kohorten im Forschungskontext sind daher vor allem allgemeine Vorgaben zur informationellen Selbstbestimmung und zum Datenschutz relevant. Beispielhaft sei auf die im Regelfall datenschutzrechtlich erforderliche Einwilligung der Betroffenen zur Erhebung, Speicherung und Verarbeitung ihrer Daten hingewiesen. Die Empfehlungen der TMF formulieren allgemeine Vorgaben für Register und Kohorten [6]. Derzeit wandeln sich Register und Kohorten von abgeschlossenen, national begrenzten Projekten zu internationalen, verteilten Plattformen. Hierdurch entstehen neue Herausforderungen, von der IT-Infrastruktur bis zur Rechtssicherheit.

Empfehlungen zur Entwicklung und zum Betrieb von Registern liegen für die USA durch die AHRQ [1] und für Deutschland durch das DNVF [3] vor. Das DNVF führt regelmäßig im Rahmen der DNVF-Spring-School Seminare zu Registern sowohl für Einsteiger als auch für Fortgeschrittene durch.

## Kategorisierung der IT-Landschaft

Für spezifische, rechnergestützte Anwendungssysteme zur Unterstützung von Registern und Kohorten kann ein sich entwickelnder Markt beobachtet werden. Im Wesentlichen sind es vier Kernprozesse, für die der Einsatz spezifischer Anwendungssysteme in Frage kommt:

- ▶ die Kommunikation mit den Studienzentren,
- ▶ die Verwaltung der Probanden,
- ▶ das Management der Daten von der Erhebung bis zum Monitoring sowie
- ▶ die Auswertung und Nutzung der Daten.

Weiterhin können Prozesse aus der Entwicklungs- oder auch Abschlussphase eines Registers oder einer Kohorte unterstützt werden, wie z. B. Planungs- oder Archivierungsaufgaben. Für die ersten beiden Kernprozesse kann an die Nutzung von so genannten Customer-Relationship-Management-Systemen (CRM) gedacht werden, wie man sie in Unternehmen der freien Wirtschaft zur Verwaltung von Kundenbeziehungen

einsetzt [7]. Für die insbesondere bei Kohorten wichtige Verwaltung der Probanden werden bislang vorwiegend Eigenentwicklungen genutzt. Für das Management von Daten dominieren Studienmanagementsysteme bzw. Systeme zum Electronic Data Capture (EDC) aus dem Umfeld klinischer Studien. Diese Systeme stoßen jedoch insbesondere bei der Unterstützung versorgungsnaher Register an ihre Grenzen, da dort die Datenverwaltung eine hohe Ähnlichkeit mit der elektronischen Akte aufweist. Nicht zuletzt werden auch allgemeine Tools für Teilaspekte des Datenmanagements eingesetzt, sei es ein Datenbankverwaltungssystem, eine Tabellenkalkulation oder ein Statistikpaket. Für den Kernprozess der Auswertung und Nutzung der Daten – bis hin zur Publikation – sind den Autoren keine rechnergestützten Anwendungssysteme bekannt, die eine spezifische Unterstützung für Register und Kohorten oder Data Repositories anbieten. Derzeit findet sich hier eine Kombination aus Werkzeugen für Teilaufgaben, wie Statistikpakete für die statistische Auswertung sowie Eigenentwicklungen z. B. für die regelmäßige Veröffentlichung von Berichten im Web. Angesichts dieser schwierigen Marktlage werden offene Baukästen entwickelt, aus denen sich Register die für sie geeigneten rechnergestützten Werkzeuge auswählen können [8]. Für den Bereich klinischer Krebsregister decken einzelne Lösungen zumindest Teilaspekte aller vier der oben genannten Kernprozesse ab. Bekanntestes Beispiel einer solch umfassenden Lösung ist das Gießener Tumordokumentationssystem (GTDS), welches seit den 90er Jahren von der Medizinischen Informatik an der Universität Gießen entwickelt und betreut wird [9].

## Anforderungen an Kohortenstudien & Register IT (KoRegIT)

Im TMF-Projekt V078-01 „Von der Evaluierung zur Konsolidierung: Anforderungen an Kohortenstudien & Register IT (KoRegIT)“ wurde unter Leitung des Zentrums für Klinische Studien Köln (ZKS Köln) ein Anforderungskatalog finalisiert [10].<sup>5</sup> Dieser ermöglicht es, die Systemlandschaft von Kohorten und Registern in ihrem Bestand und hinsichtlich des Bedarfs zu beschreiben und zu vergleichen. Der entwickelte Anforderungskatalog beschreibt für die oben genannten Kernprozesse alle typischen oder zumindest bedenkenswerten Aufgaben, die eine Kohorte oder ein Register zu erfüllen hat und die mindestens perspektivisch von einer IT-Lösung unterstützt werden können. Der Anforderungskatalog ist über den Lebenszyklus eines Vorhabens gegliedert und weist in den folgenden Phasen Top-Level-Aufgaben und Anwendungsfälle (Use Cases) aus:

- ▶ Entwurf: Planung, Vorbereitung
- ▶ Betrieb: Umsetzung, Nutzung, Weiterentwicklung
- ▶ Abschluss

---

<sup>5</sup> Download unter <http://www.tmf-ev.de/Produkte/P100901>.

Die Systematik des Anforderungskatalogs über Top-Level-Aufgaben und Use Cases wurde zunächst ausgehend von den Kompetenznetzen HIV/AIDS und Hepatitis festgelegt. Der daraufhin erarbeitete Anforderungskatalog wurde mit verschiedenartigen Vorhaben wie z. B. der NAKO Gesundheitsstudie (damals „Nationale Kohorte“), dem Mukoviszidose-Register sowie Klinischen Krebsregistern abgestimmt und weiterentwickelt. Abschließend wurden Ergebnisse aus vier Experten-Reviews eingearbeitet.

In aktueller Fassung werden über die verschiedenen Phasen 44 Top-Level-Aufgaben und 203 Use Cases definiert. Die Use Cases enthalten jeweils eine kurze Beschreibung der Aufgabe, benennen die notwendigen bzw. beteiligten Akteure sowie die Vor- und Nachbedingungen und skizzieren einen Standardablauf. Anhand der Vorbedingungen werden die Abhängigkeiten zwischen den Anwendungsfällen ersichtlich, aus denen sich z. B. für die Planung eines Vorhabens Ablaufpläne ableiten lassen. Um die Use Cases einheitlich zu beschreiben, wurden die beteiligten Akteure gemäß ihrer Funktion und Rolle definiert. Die Akteure wurden zudem gruppiert in „Akteure der teilnehmenden Zentren“, „Akteure der zentralen Struktur“ sowie „Kooperierende Akteure“. Weiterhin sind Komitees und Gremien benannt, die sich aus den einzelnen Akteuren zusammensetzen. Diese Übersicht kann an ein konkretes Vorhaben angepasst und entsprechend zur Personal- und Kostenplanung für Kohorten und Register herangezogen werden.

Der Anforderungskatalog besteht aus zwei Dokumenten. Das erste Dokument führt die Top-Level-Aufgaben und Use Cases auf; das zweite Dokument enthält eine detaillierte Spezifikation, wobei alle Use Cases gemäß der oben genannten Vorlage ausgearbeitet sind. Die Liste der Top-Level-Aufgaben und Use Cases kann auch als Checkliste für eine erste Planung eines Registers oder einer Kohorte genutzt werden. Tabelle 4.1 zeigt alle Top-Level-Aufgaben in den oben genannten Phasen mit ihren eindeutigen Identifikationsnummern. Der Anforderungskatalog aus KoRegIT kann entsprechenden Projekten in jeder Phase ihres Lebenszyklus Anleitung und Unterstützung in Bezug auf die einzusetzende oder eingesetzte IT-Infrastruktur bieten. Zudem kann der allgemeine Aufbau und die Organisationsstruktur von Kohorten und Registern abgeleitet werden.

Die Nutzung des KoRegIT-Katalogs zur Planung einer IT-Infrastruktur ist in zwei Richtungen denkbar. Zum einen kann auf funktionale Zusammenhänge zurückgegriffen werden, die durch ein rechnergestütztes System zu unterstützen sind, z. B. auf Ebene der Top-Level-Aufgaben. Zum anderen kann über Phasen, Top-Level-Aufgaben und Use Cases hinweg eine einheitliche Unterstützung am Rechnerarbeitsplatz für Mitarbeiter in ihren spezifischen Rollen angestrebt werden.

## Tabelle 4.1: Top-Level-Aufgaben in KoRegIT

Phase	ID	Bezeichnung
Entwicklung/Planung	TL002	Entwicklung und Festlegung der wissenschaftlichen Fragestellungen
	TL003	Protokollentwicklung
	TL001	Planung der Organisationsstrukturen
	TL004	Gewinnung von und Abstimmung mit Kooperationspartnern
	TL047	Finanzplanung
	TL006	Erstellung Probandeninformation und Einwilligungserklärung
	TL005	Erstellung und Abstimmung Datenschutzkonzept
	TL041	Prüfung Versicherungspflicht/Abschluss Versicherungen
	TL007	Erstellung und Einreichung Ethikantrag/Behördenmeldung
	TL044	Erstellung und Abstimmung Qualitätsmanagementkonzept
Entwicklung/Vorbereitung	TL008	Festlegung des Meldeverfahrens
	TL009	Erstellung eines statistischen Analyseplans
	TL042	Planung des Berichtswesens
	TL010	Vorbereitung von Datenerhebung und Datenerfassung
	TL048	Erstellung und Abstimmung Monitoring-Manual
	TL011	Erstellung und Pflege von Verfahrensanweisungen
	TL012	Erstellung von Dokumentvorlagen
	TL013	Erstellung von Vertragsvorlagen
	TL014	Aufbau der Organisationsstruktur
	TL015	Rekrutierung von Zentren
	TL039	Erstellung und Umsetzung des Berechtigungskonzepts
	TL043	Schulung des Zentrumspersonals
	TL040	Dienstleistungen definieren und Abrechnungssystem implementieren
	TL018	Projektorganisation und Ressourcenplanung
	TL019	Nutzer- und Stammdatenverwaltung
TL027	Zugriff auf Wissen	
Betrieb/Umsetzung	TL022	Probandenmanagement
	TL037	Unterstützung und Betreuung von Probanden
	TL021	Datenerhebung und Datenerfassung
	TL023	Monitoring
	TL024	Bereitstellung von probandenbezogenen Informationen
	TL025	Abrechnung mit Erhebungszentren und Probanden
	TL045	Abrechnung von Dienstleistungen
	TL026	Datenmanagement (Organisation und Pflege der Daten)

**Tabelle 4.1: Top-Level-Aufgaben in KoRegIT (Fortsetzung)**

Phase	ID	Bezeichnung
Betrieb/Nutzung	TL028	Studienunterstützung
	TL029	Statistische Analyse
	TL030	Berichterstattung
	TL031	Organisation von Publikationen und Präsentationen
	TL033	Datenintegration, Datenzusammenführung
	TL046	Unterstützung der Patientenversorgung
Betrieb/Weiterentwicklung	TL032	Weiterentwicklung Register/Kohorte
Abschluss	TL034	Archivierung
	TL035	Vernichtung der Daten, Anonymisierung
	TL036	Close Out

Die Reihung der Top-Level-Aufgaben folgt positiven Erfahrungen der Autoren des Katalogs über eine sinnvolle, schrittweise Entwicklung und Inbetriebnahme eines Registers oder einer Kohorte. Die Reihenfolge ist selbst allerdings nicht Gegenstand der Empfehlung. Der numerische Anteil der eindeutigen IDs reflektiert Eigenschaften des Projektverlaufs und hat darüber hinaus keine Bedeutung.

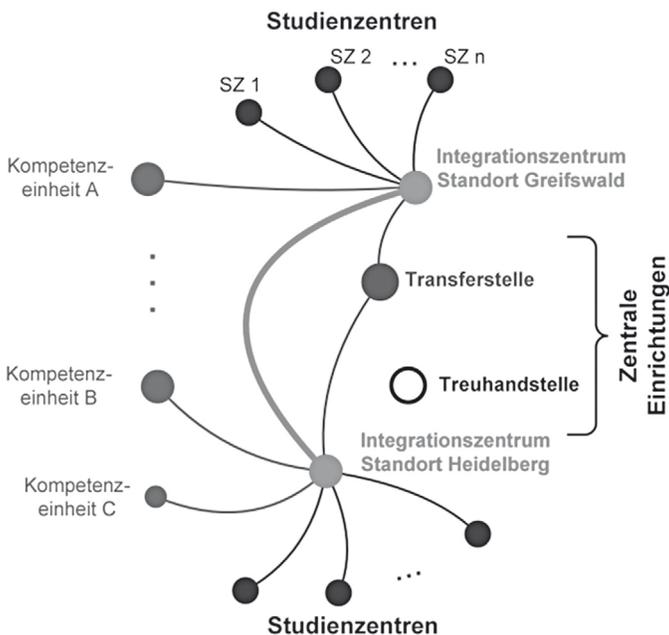


Abbildung 4.1: Infrastruktur der NAKO Gesundheitsstudie.

# IT-Infrastruktur der NAKO

## Gesundheitsstudie

Die NAKO gilt als größte Gesundheitsstudie in Deutschland. Die Baseline-Erhebungen der geplanten 200.000 Teilnehmer erfolgen in 18 Studienzentren und umfassen Befragungen, Untersuchungen sowie die Einlagerung von Bioproben. Geplant sind die Verknüpfung mit Sekundärdaten sowie Nachbeobachtungen und Folgeuntersuchungen.

Besondere Herausforderungen für die IT-Infrastruktur ergeben sich aus der dezentralen Organisation, dem zeitlichen und inhaltlichen Umfang der Datenerhebung sowie Datenschutz- und ethischen Anforderungen. Hohe Effizienz und Datenqualität sind nur bei weitgehender Standardisierung, Zentralisierung und Automatisierung aller Datenverarbeitungsprozesse erreichbar. Dazu wurde ein Datenmanagementkonzept entwickelt, das neben den jeweils in Teilnehmermanagement und Untersuchungszentrum gegliederten Studienzentren eine unabhängige Treuhandstelle zur zentralen Verwaltung von personenidentifizierenden Daten, Einwilligungen, Widerrufen und Pseudonymzuordnungen, ein auf zwei Standorte verteiltes Integrationszentrum zur Zusammenführung der dezentral erhobenen Studiendaten, fachspezifische Kompetenzeinheiten zur standardisierten Aufbereitung komplexer Rohdaten und eine Transferstelle zur Bereitstellung der Studiendaten und Bioproben umfasst (siehe Abbildung 4.1).

Die pseudonymisierte Datenerhebung in den Untersuchungszentren erfolgt standardisiert mit einer vom Integrationszentrum bereitgestellten Webanwendung. Neben der Dokumentation von Untersuchungsabläufen und -ergebnissen beinhaltet diese Schnittstellen zum Laborinformationsmanagementsystem und zu Zentralsystemen der MRT-Untersuchung und ermöglicht sowohl die Erstellung von Ergebnismitteilungen an die Teilnehmer als auch die Abrechnung der Probandenpauschalen.

Zentrales Element für alle Phasen des Datenmanagements ist das Data Dictionary. Dieses umfasst eine Beschreibung aller Variablen und Untersuchungen in einem einheitlichen Format inklusive der Validierungs- und Sprungregeln sowie der Schritte der Prä- und Postprozessierung.

In die Webanwendung zur Datenerhebung integriert ist die Erfassung von Einwilligungs- und Schweigepflichtentbindungserklärungen. Diese werden online auf biometrischen Signpads unterschrieben. Da hier identifizierende Daten verarbeitet werden, erfolgt die Datenübertragung unmittelbar zur Treuhandstelle. Vor jeder Verarbeitung personenbezogener Daten wird automatisiert das Vorliegen der dazu erforderlichen Einwilligung geprüft.

Über ein Java-Programm auf dem jeweiligen Untersucher-PC werden die Initialisierung von Untersuchungen aus bereits in der Webanwendung erhobenen Studiendaten sowie Upload und Auswertung von Ergebnisdaten diagnostischer Geräte gesteuert. Weitere Bestandteile der IT-Infrastruktur sind ein Linux-basiertes Kiosk-Betriebssystem für die Bearbeitung von Selbstausfüller-Fragebögen am Touchscreen, Open-Source-Tools zum Software-Versionmanagement, zum Monitoring und zum Bugtracking sowie eine Software zur Kontakt- und Terminverwaltung im Teilnehmermanagement.

## Aktueller Bestand

### Einleitung

Eine systematische Übersicht zu Registern und Kohorten existiert in Deutschland derzeit nicht bzw. nur für ausgewählte Anwendungsgebiete. Informationen zu bevölkerungsbezogenen Krebsregistern sind über die Gesellschaft der epidemiologischen Krebsregister (GEKID), Informationen zu klinischen Krebsregistern über die Arbeitsgemeinschaft Deutscher Tumorzentren (ADT) verfügbar. Eine Übersicht zu Registern im Bereich der Medizintechnik wurde publiziert [11]. Für seltene Erkrankungen besteht über Orphanet eine Recherchemöglichkeit unter Einschluss deutscher Vorhaben.<sup>6</sup> Hinweise auf große Kohorten lassen sich den Unterlagen der NAKO Gesundheitsstudie entnehmen [12].

Durch die insgesamt fehlende Transparenz zu bestehenden Registern und Kohorten entsteht nicht nur ein Informationsdefizit, sondern ebenso ein Qualitätsproblem. Unter anderem müssen Mehrfachentwicklungen für überlappende Versorgungssituationen, Unkenntnis von fachlichen Standards bei der Entwicklung und dem Betrieb dieser Vorhaben sowie Rekrutierungsprobleme als Folge der mangelnden Transparenz erwartet werden. Vergleichbar dem Deutschen Register für Klinische Studien (DRKS) und dem Deutschen Biobanken-Register (DBR) haben daher DNVF und TMF eine Initi-

<sup>6</sup> Siehe <http://www.orpha.net/> oder <http://www.orpha.net/national/DE-DE/>.

ative zur Implementierung eines Registerportals – eines Registers von Registern und Kohorten – gestartet [13]. Eine erste Datengrundlage hierfür sind die Ergebnisse aus Erhebungen zu Registern der Mitgliedsverbände der TMF. Die Ergebnisse der letzten Erhebung von 2012 werden im Folgenden vorgestellt.

## **Erhebung der TMF von 2012**

Erstmals wurden im November 2007 alle Mitgliedsverbände der TMF um Informationen über ihre Register gebeten. Aus elf von 55 Verbänden wurden damals Angaben zu 14 Registern erfasst [14]. Im Jahre 2009 erfolgte dann im Rahmen von Überlegungen zur IT-Strategie der TMF eine Analyse der Anforderungen und des Ausstattungsstandes in Bezug auf die IT medizinischer Forschungsregister [15]. Zehn Betreiber von Registern wurden im Rahmen halb-standardisierter Interviews zu Konzept und Umsetzung des Registers, zu technischen Aspekten, zu Aufgaben des zentralen Registerbetreibers im Servicemanagement und zur Definition von Registern befragt. Die 2009 gewonnenen Ergebnisse wurden mit der AG Register des DNVF diskutiert und sind in das vom DNVF herausgegebene „Memorandum Register für die Versorgungsforschung“ eingegangen [3]. Ausgehend von den beiden vorangehenden Projekten wurde für die Erhebung zu Registern und Kohorten von 2012 ein Erhebungsbogen erarbeitet und mit dem IT-Reviewing-Board abgestimmt. Primär wurde nach dem Betrieb von Registern gefragt. Kohorten konnten ebenfalls gemeldet werden. Eine fehlende Rückmeldung wurde als Nichtvorhandensein eines Registers interpretiert.

Insgesamt konnten Angaben zu 31 Registern und Kohorten von 27 Verbänden gewonnen werden. Damit betrieben etwa 30% der eingeschlossenen 88 Verbände ein Register oder eine Kohorte. Das erste Register wurde 1995 in Betrieb genommen, zwischen 1995 und 2006 14 Register und zwischen 2008 und 2012 16 Register. Bei einem Register fehlte diese Information. Bei 26 Registern wurden die zentralen IT-Komponenten von Einrichtungen im universitären Umfeld betrieben. Bei fünf Registern fanden sich andere Betreibermodelle.

Die Register verfügten im Median über eine Vollkraftstelle als zentrale Ressource. Zehn Register gaben zusätzlich Studienpersonal in den Erhebungszentren an. Die Zahl der Erhebungszentren lag zwischen 1 und 194; im Median betrug sie 11. Folgende Einrichtungen waren dabei als Erhebungszentrum beteiligt: Universitätsklinik – 29 Register, nicht-universitäre Krankenhäuser – 20, niedergelassene Ärzte – 15, Probanden/Patienten – 7, Angehörige – 6, Sonstige – 6. Die geplante Laufzeit wurde von 15 Registern mit „unbegrenzt“ angegeben, bei weiteren 11 Registern lag diese bei mindestens 10 Jahren. Entsprechend war bei den meisten Registern auch ein unbegrenzter

Meldezeitraum vorgesehen. Bei 14 Registern war die Zahl der Teilnehmer unbegrenzt bzw. wurde eine komplette Erhebung der Zielpopulation angestrebt. Von 27 Registern lagen Angaben zur Zahl aktuell registrierter Patienten vor. Im Median lag diese Zahl bei 710 (Minimum 16, Maximum 143.000, Q1 287, Q3 8.728). Die Verteilung zeigt Abbildung 4.2.

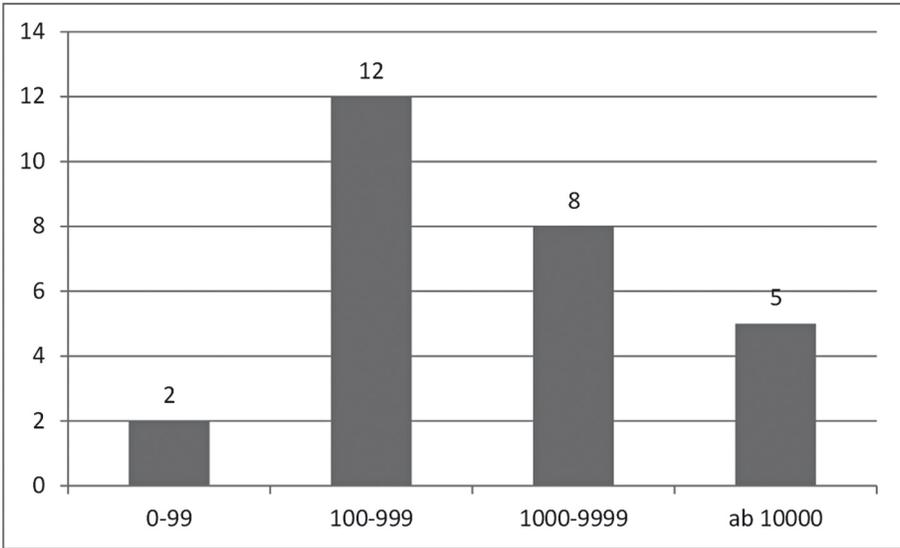


Abbildung 4.2: Zahl der Register nach Anzahl registrierter Patienten oder Probanden.

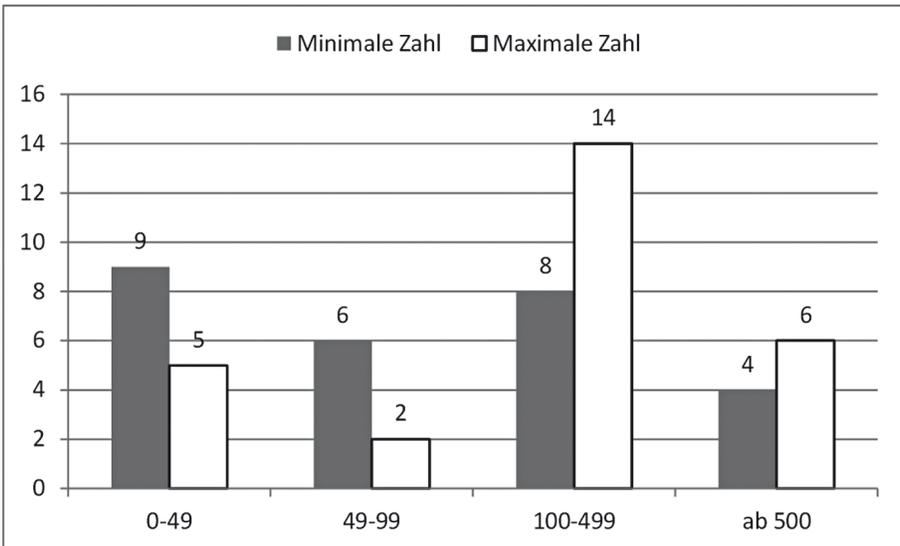


Abbildung 4.3: Anzahl der Register nach minimaler und maximaler Zahl der zu erfassenden Merkmale.

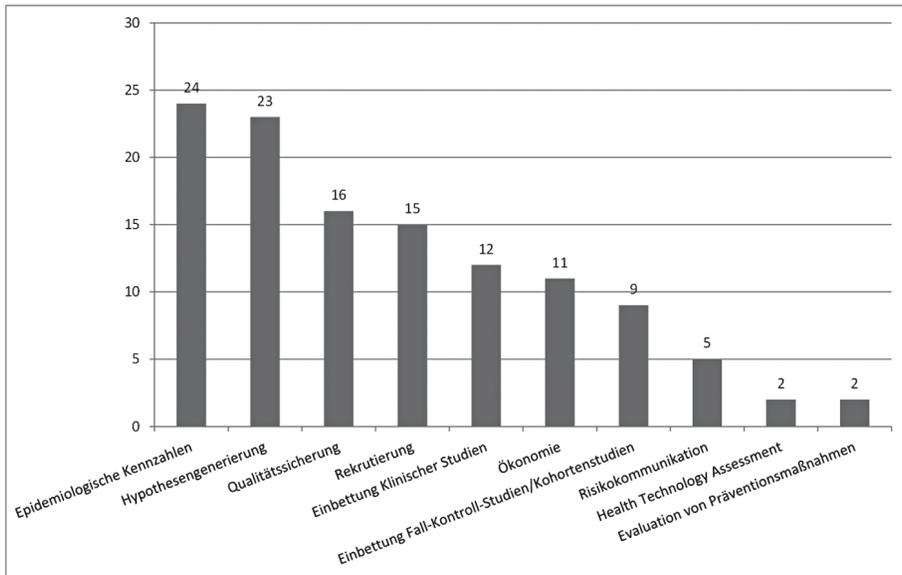


Abbildung 4.4: Von den Registern erreichte Ziele.

Die Mindestzahl der pro Proband zu erfassenden Merkmale lag zwischen 5 und 2.700. Die Verteilung ist in Abbildung 4.3 dargestellt. Ein Follow-up führten 28 Register durch; drei Register verzichteten auf ein Follow-up. Siebzehn Register führten das Follow-up lebenslang durch. Bei den Betriebskonzepten dominierte ein zumindest teilweises Outsourcing mit 11 Nennungen, gefolgt von einem Betrieb von zumindest Teilen der IT-Infrastruktur durch ein Universitätsklinikum bei 8 Registern. Mit vier Nennungen war ein kompletter Eigenbetrieb seltener.

Dreiundzwanzig Register stellten den Mitgliedern des Verbunds, 14 der Öffentlichkeit regelmäßig Analysen zur Verfügung. Bei 15 Registern erhielten die Erhebungszentren die erfassten Daten zurück. Ebenfalls 15 Register gaben die Daten an Dritte weiter. Eine Übersicht zur Erreichung der Ziele enthält Abbildung 4.4.

## Verfügbare Softwareprodukte

### Softwareprodukte zur Unterstützung von Kernprozessen

In einem Workshop „Register-Toolbox“ der TMF wurden Anfang 2014 kommerzielle und nicht-kommerzielle Softwareprodukte für den Einsatz bei Registern und Kohorten vorgestellt<sup>7</sup> Schon damals bestanden Überlegungen, eine Marktübersicht zu schaffen.

<sup>7</sup> siehe <http://www.tmf-ev.de/news/1461>

Ende Oktober 2015 wurden dann Anbieter und Experten um Informationen zu 24 Produkten gebeten. Hierbei handelte es sich um Produkte aus dem Workshop von 2014, aus anderweitigen Kontakten der TMF sowie aus einer diesbezüglichen Anfrage an die Mitglieder der AG IT-Infrastruktur und Qualitätsmanagement der TMF. Als Voraussetzung für den Einschluss einer Software wurden vier Bedingungen genannt: 1) die Software wird spezifisch für Register und/oder Kohorten angeboten, 2) die Software deckt für Register und Kohorten mindestens den Prozess des Datenmanagements ab, 3) ein Beschaffungsweg zum Bezug der Software ist etabliert und 4) es besteht mindestens eine vom Entwicklungsstandort unabhängige Nutzung. Neben einem dreiseitigen Fragebogen umfasste die Erhebung den Katalog der Top-Level-Aufgaben aus KoRegIT, um den aktuellen und geplanten Funktionsumfang der Software zu erfassen. Zu 15 Produkten oder Produktgruppen lagen bis zum damaligen Redaktionsschluss Rückmeldungen vor, davon umfassten zwei Rückmeldungen Werkzeuge jenseits des Datenmanagements. Die verbleibenden 13 Produkte wurden in die Übersicht aufgenommen:

- ▶ Adjumed.net, Adjumed Services AG (<http://www.adjumed.net/>)
- ▶ BQS-Register, BQS Institut für Qualität und Patientensicherheit GmbH (<http://www.bqs-institut.de/>)
- ▶ BRT - BAP Register Technologie, BioArtProducts GmbH (<http://bioeng.de/>)
- ▶ MARVIN, XClinical GmbH (s. <http://www.xclinical.com/>)
- ▶ OpenClinica, OpenClinica, LLC (<https://openclinica.com/>)
- ▶ OSSE (Open-Source-Registersystem für Seltene Erkrankungen in der EU), Medical Informatics Group, Universitätsklinikum Frankfurt ([http://www.osse-register.de](http://www.osse-register.de/))
- ▶ ProMISe (Project Manager Internet Server), Leiden University Medical Center (<http://www.msbi.nl/promise>)<sup>8</sup>
- ▶ REDCap, Vanderbilt University (<http://project-redcap.org/>)
- ▶ secuTrial, interActive Systems GmbH (<http://www.secutrial.com/>)
- ▶ Smart Reporting, QMedify GmbH (<http://www.smartreporting.de/>)
- ▶ triplextrax, axaris GmbH (<http://www.axaris.de/>)
- ▶ VRI - Visionet Register Infrastruktur, Visionet GmbH (<http://www.visionet.de>)
- ▶ WebSpirit Clinical Trial Management System (CTMS), 2mt Software GmbH (<http://www.2mt-software.de/>)

Tabelle 4.2 stellt allgemeine Angaben zu den Produkten zusammen. Die Abbildung der Funktionalität auf die Top-Level-Aufgaben von KoRegIT zeigt eine klare Fokussierung des Software-Angebots im Sinne eines Electronic Data Capture (s. Tabelle

---

<sup>8</sup> Zu ProMISe wurden keine Informationen zum Funktionsumfang eingereicht, das Werkzeug wurde jedoch im Rahmen des genannten Workshops vorgestellt.

4.3). Die Datenerhebung wird hinsichtlich der Merkmalsdefinition vorbereitet; die Datenerfassung erfolgt über die Software, die dann auch das Datenmanagement z. B. mit Plausibilitätsprüfungen unterstützt. Dazu muss für die Studienzentren und die Anwender ein Berechtigungskonzept mit einer Nutzer- und Stammdatenverwaltung umgesetzt sein. Datenintegration und Archivierung schließen den Prozess des Datenmanagements ab. Eine Unterstützung der Entwicklung von Registern und Kohorten in Planung und Vorbereitung sowie eine Unterstützung organisatorischer Aufgaben wird hingegen seltener angegeben. Die Kommunikation mit Studienzentren ist in der Breite Einsatzgebiet von eigenständigen Softwareprodukten, während eine Unterstützung des Probandenmanagements vielfach bestätigt wird. Von den Phasen des Lebenszyklus eines Registers oder einer Kohorte werden vor allem jene des Betriebs unterstützt und hier am häufigsten die Umsetzung, gefolgt von der Nutzung. Dem Schwerpunkt auf dem Betrieb steht eine geringere Unterstützung in der Vorbereitung sowie insbesondere bei der Planung gegenüber.

**Tabelle 4.2: Software für Register und Kohorten (Selbstauskunft)**

Software- produkt	Nutzung			Prozesse <sup>1</sup>					
	Beginn	Anzahl	Lizenz <sup>2</sup>	DM	E	PM	KOM	AUS	ARCH
Adjumed.net	2007	15	kommerziell	voll	nein	weit.	nein	voll	voll
BQS-Register	2010	9	kommerziell	voll	voll	voll	voll	voll	voll
BRT	2012	4	kommerziell	voll	teil	voll	voll	voll	voll
MARVIN	2003	> 25	kommerziell	voll	voll	voll	voll	nein	voll
OpenClinica <sup>3</sup>	2006		LGPL	voll	teil	weit	weit	teil	teil
OSSE	2012	nicht bekannt	AGPLv3	voll	teil	weit	teil	teil	nein
ProMISe	1995	250	ASP	voll	voll	voll	voll	voll	teil
REDCap <sup>3</sup>	2008		kostenfrei, keine Open- Source-Lizenz	voll	teil	teil	nein	teil	teil
secuTrial	2000	25	kommerziell	voll	teil	weit	voll	voll	voll
Smart Reporting	2016	1	kommerziell	voll	teil	teil	teil	teil	teil
tripletrax	2015	1	kommerziell	voll	teil	teil	voll	teil	voll
VRI	2007	2	kommerziell	teil	teil	weit	teil	teil	weit
WebSpirit	2004	25	kommerziell	voll	teil	weit	teil	teil	weit

<sup>1</sup> DM = Datenmanagement, E = Entwicklung, PM = Probandenmanagement, KOM = Kommunikation mit Studienzentren, AUS = Datenanalyse oder -verwertung, ARCH = Archivierung

Kategorien der Prozessunterstützung: voll = vollständig, teil = teilweise, nein, weit. = über weitere Module oder weitere Software

<sup>2</sup> ASP = Application Service Providing, LGPL = GNU Lesser General Public License, AGPL = GNU Affero General Public License

<sup>3</sup> Die Informationen zu OpenClinica und RedCap wurden von deutschen Anwendern eingebracht. Da die zugänglichen Nutzungszahlen u. a. klinische Studien umfassen, wurde auf deren Wiedergabe verzichtet.

**Tabelle 4.3: Abdeckung von Top-Level-Aufgaben durch das Softwareangebot (Selbstauskunft 2015)**

Phase	KoReGIT Top-Level-Aufgabe	Anzahl Software-Produkte (Gesamtzahl: 12)					fehlende Angabe
		produktiv genutzt	vorhanden	in Entwicklung	geplant	nicht geplant	
Entwicklung/ Planung	Entwicklung und Festlegung der wissenschaftlichen Fragestellungen	3				9	
	Protokollentwicklung	3				9	
	Planung der Organisationsstrukturen	2			1	9	
	Gewinnung von und Abstimmung mit Kooperationspartnern	1	1		1	9	
	Finanzplanung	1			1	10	
	Erstellung Probandeninformation und Einwilligungserklärung	2	1	1		8	
	Erstellung und Abstimmung Datenschutzkonzept	4	1		1	6	1
	Prüfung Versicherungspflicht/Abschluss Versicherungen	1	1			9	
	Erstellung und Einreichung Ethikantrag/Behördenmeldung	2	1		1	9	
	Erstellung und Abstimmung Qualitätsmanagementkonzept	2		1		8	
Entwicklung/ Vorbereitung	Festlegung des Meldeverfahrens	1	1		1	10	
	Erstellung eines statistischen Analyseplans	2	1	1	1	8	
	Planung des Berichtswesens	3	1		1	6	
	Vorbereitung von Datenerhebung und Datenerfassung	9				2	
	Erstellung und Abstimmung Monitoring-Manual	2			2	8	
	Erstellung und Pflege von Verfahrensanweisungen	2			2	8	
	Erstellung von Dokumentvorlagen	4			1	7	
	Erstellung von Vertragsvorlagen	2			1	9	
	Aufbau der Organisationsstruktur	4	1			8	
	Rekrutierung von Zentren	9	2		2	9	
	Erstellung und Umsetzung des Berechtigungskonzepts	5	1			1	
	Schulung des Zentrumspersonals	3				6	
	Dienstleistungen definieren und Abrechnungssystem implementieren	2				9	

Tabelle 4.3: Abdeckung von Top-Level-Aufgaben durch das Softwareangebot (Selbstauskunft 2015) (Fortsetzung)

Phase	KoReGIT Top-Level-Aufgabe	Anzahl Software-Produkte (Gesamtzahl=12)					fehlende Angabe
		produktiv genutzt	vorhanden	in Entwicklung	geplant	nicht geplant	
Entwicklung/ Vorbereitung	Projektorganisation und Ressourcenplanung	10	1	1		10	
	Nutzer- und Stammdatenverwaltung Zugriff auf Wissen	5			1	6	
Betrieb/Um- setzung	Probandenmanagement	7	1		2	2	
	Unterstützung und Betreuung von Probanden	3	1		1	5	2
	Datenerhebung und Datenerfassung	11	1				
	Monitoring	9	1	1		1	
	Bereitstellung von probandenbezogenen Informationen	4		1	1	6	
	Abrechnung mit Erhebungszentren und Probanden	3	1		1	7	
	Abrechnung von Dienstleistungen	3			2	7	
	Datenmanagement (Organisation und Pflege der Daten)	11	1				
Betrieb/ Nutzung	Studienunterstützung	6	1	1		4	
	Statistische Analyse	7	2			3	
	Berichterstattung	6	1	3		2	
	Organisation von Publikationen und Präsentationen	3			1	8	
	Datenintegration, Datenzusammenführung	8	1	1		2	
	Unterstützung der Patientenversorgung	3			1	8	
Betrieb/Wei- terentwick- lung	Weiterentwicklung Register/Kohorte	8	2	1	1		
Abschluss	Archivierung	8				4	
	Vernichtung der Daten, Anonymisierung	7	2		1	2	
	Close Out	7		1	1	3	

Die Top-Level-Aufgaben beziehen sich auf den im KoRegIT-Projekt der TMF erstellten Anforderungskatalog zur IT-Unterstützung von Kohorten und Registern. Weitere Erläuterungen siehe Text.

## Softwareprodukte zur Unterstützung weiterer Aufgaben

Am Oldenburger OFFIS-Institut für Informatik wurde die Software Suite CARTools für das Niedersächsische Krebsregister entwickelt, bestehend aus den Anwendungen CARELIS und CARESS.<sup>9</sup> CARELIS ist ein Werkzeug zur Zusammenführung von pseudonymisierten Datensätzen im Sinne eines Privacy Preserving Record Linkage. CARESS ist ein Data Warehouse System zur Auswertung und statistischen Analyse von Daten aus Krebsregistern [16]. Die explorative CARESS Auswertungssoftware liefert krebsepidemiologische Kennzahlen für Routineauswertungen. Die Ergebnisse können über verschiedene Schnittstellen abgefragt und auf vielfältige Art und Weise visualisiert werden. Über eine Schnittstelle zur Statistik-Umgebung R können Anwender weitere statistische Auswertungen vornehmen. Für die interaktive Berichterstattung stehen Vorlagen zur Verfügung.

Das DFG-geförderte Projekt MOSAIC stellt zahlreiches Material bereit, um praktische Hilfestellung bei der Realisierung eines Registers oder einer Kohorte zu bieten [17].<sup>10</sup> Zum Identitätsmanagement wird das Enterprise Person Identifier Crossreferencing (E-PIX) angeboten. E-PIX setzt das Konzept eines Master Patient Index um und bietet die technische Funktionalität zur eindeutigen, standortübergreifenden Identifizierung von Personen. Über den Generic Pseudonym Administration Service (gPAS) erfolgt eine Pseudonymgenerierung und -verwaltung. Das Konzept unterstützt unterschiedliche Pseudonyme je Datenquelle, Anwendungskontext und Standort. Der Generic Informed Consent Service (gICS) erlaubt die Verwaltung modularer Einwilligungen und Widerrufe. Der Einwilligungsstatus eines Datensatzes kann automatisierbar geprüft und die Dokumentation des Einwilligungsprozesses unterstützt werden. mosaicQA unterstützt Forscher in der Erstellung grundlegender Qualitätsberichte für metrische und kategoriale Daten mit R. mosaicQA wird als offizielles R-Package bereitgestellt. Die MOSAIC Toolbox for Research stellt eine Möglichkeit zur web-basierten und standortübergreifenden Datenerfassung für kleinere Vorhaben unter Verwendung von OpenClinica dar. Weitere Materialien aus MOSAIC unterstützen die Erstellung eines Datenschutzkonzepts (Vorlage), die Konzeptionierung von Datensicherung und -wiederherstellung (Vorlagen) sowie die Erstellung elektronischer Erhebungsformulare und eines Data Dictionary (Leitfäden).

Bei dem in Tabelle 4.2 bereits vorgestellten Open-Source-Registersystem für Seltene Erkrankungen in der EU (OSSE) finden sich auch Komponenten zur Unterstützung

---

<sup>9</sup> Siehe <http://www.offis.de>.

<sup>10</sup> Siehe <http://mosaic-greifswald.de/>.

# Versorgungsforschung mit und für Menschen mit Multipler Sklerose (MS) – das Deutsche MS-Register

2001 initiierte die Deutsche Multiple Sklerose Gesellschaft, Bundesverband e.V., (DMSG) die Gründung eines Registers, um erstmals deutschlandweit Daten zu Erkrankten verfügbar zu machen. Das Register umfasst longitudinale Daten, u. a. zu Verlaufsformen, soziodemographischen Faktoren und der Versorgungssituation von Menschen mit MS in Deutschland, die versorgungsnah von medizinischem Fachpersonal und den Patienten selbst dokumentiert werden [18]. Über die Jahre konnte ein Netzwerk von mehr als 160 neurologischen Praxen sowie Akut-, Universitäts- und Reha-Kliniken etabliert werden, die Daten an das Register melden. Das Register wird von der MS Forschungs- und Projektentwicklungs-gGmbH betrieben und weiterentwickelt, einer Tochter der Deutschen MS-Stiftung.

In seiner über 10-jährigen Geschichte wurde das Deutsche MS-Register kontinuierlich weiterentwickelt und an die wachsenden Anforderungen inhaltlich wie methodisch angepasst. Von 2010 bis 2013 durchlief das MS-Register eine Revision, in deren Verlauf in Zusammenarbeit mit dem Krankheitsbezogenen Kompetenznetz Multiple Sklerose (KKNMS) der gemeinsame Minimal- und Basisdatensatz der DMSG und des KKNMS [19] definiert wurde – ein modulares Datenmodell, welches als Grundlage für die auf Versorgungsforschung ausgerichtete Registerdokumentation dient. Darüber hinaus wurde im europäischen Verbund von MS-Registern ein Minimaldatensatz abgestimmt, welcher Eingang in den überarbeiteten Registerdatensatz gefunden hat. Die Konzeption und Umsetzung der revidierten IT-Infrastruktur des Registers erfolgte durch das Institut für Medizinische Informatik der Universitätsmedizin Göttingen. Das zentrale Element für die Datenerfassung und das Datenmanagement bildet eine webbasierte Registerdatenbank (umgesetzt mit der Software secuTrial). Anhand des gemeinsamen zentrumsübergreifenden Identitätsmanagements mit dem KKNMS sind Auswertungen über die Datenbestände des Registers und des KKNMS hinweg möglich. Ausgewählte Dokumentationssysteme sind über definierte Schnittstellen angebunden, um bereits elektronisch erfasste Daten in die Registerdatenbank zu übertragen und somit den Dokumentationsaufwand in den Zentren so gering und fehlerresistent wie möglich zu halten.

Diese Schnittstellen werden weiterhin genutzt, um bestehende „Altdatenbestände“ in die neue Infrastruktur zu migrieren. Durch den modularen Aufbau des Datensatzes sind fragestellungsspezifische Erweiterungen möglich, die allen oder ausgewählten teilnehmenden Zentren bereitgestellt werden können. Im Rahmen solcher Erweiterungen können Fragestellungen mit deutlich geringerem Aufwand beantwortet werden, als in separaten spezifisch aufgesetzten Studien, da vorhandene Items aus der „Basis-Registerdokumentation“ weiterverwendet werden können. Seit 2014 werden Patienten über die Erfassung eines validierten Lebensqualitätsfragebogens in die Registerdokumentation einbezogen, welcher im Rahmen einer Pilotphase sowohl über eine Smartphone-App als auch web-basiert in der Registerdatenbank erfasst werden kann [20]. Die hierdurch gewonnenen Informationen sollen die Berücksichtigung der Patientenperspektive für wissenschaftliche Forschungsfragestellungen stärken. Zunehmend werden Forschungsprojekte unterstützt, die internationale Repositorien aufbauen und Daten übergreifend auswerten [21].

weiterer Aufgaben. Jedes OSSE-Register beinhaltet die Komponente OSSE-Data-management zum Speichern von Registerdaten (OSSE.EDC). Die in OSSE.EDC zu erfassenden Datenelemente werden in einem Metadaten-Verzeichnis (MDR) spezifiziert. OSSE.EDC kann zusammen mit dem MDR alleine betrieben, das MDR auch ohne OSSE.EDC genutzt werden. Für das Identity-Management wird die sogenannte Mainzliste<sup>11</sup> mitgeliefert (s. Kapitel zum Identitätsmanagement). Über den OSSE-Brückenkopf können andere Register an OSSE angebunden werden. Voraussetzung ist dabei die Abbildung der in diesen Registern befindlichen Daten im MDR. Samplify.Auth stellt einen Authentifizierungsdienst zur Verfügung. Weitere Materialien unterstützen die Erstellung eines Datenschutzkonzepts (Datenschutz-Schablone) sowie einer Patienteneinwilligung (Patienteneinwilligung-Schablone).

## Epidemiologische und klinische Krebsregister

### Überblick

Im Bereich der Krebsregister findet derzeit eine tiefgreifende Neustrukturierung statt. Das Krebsfrüherkennungs- und -registergesetz (KFRG) vom April 2013 bildet die Grundlage für den flächendeckenden Aufbau einer klinischen Krebsregistrierung. Während Ende der 90er Jahre zunächst eine flächendeckende epidemiologische Krebsregistrierung etabliert wurde, gab es flächendeckende regionale klinische Krebsregister

<sup>11</sup> Siehe <http://www.mainzliste.de/>.

bislang nur in den östlichen, d. h. den neuen Bundesländern (außer Berlin) und Bayern. Die klinische Krebsregistrierung in den westlichen Bundesländern und Berlin war einrichtungsbezogen auf Kliniken oder Klinikverbände begrenzt und erlaubte hier keine Aussagen über die Versorgung in der Region. Im Rahmen des 2008 von Bundesministerium für Gesundheit, Deutscher Krebsgesellschaft, Deutscher Krebshilfe und ADT initiierten Nationalen Krebsplans wurde als eine Maßnahme der erwähnte flächendeckende Ausbau klinischer Krebsregistrierung gefordert, um dieses Defizit zu beheben.

Um zu verstehen, was dies für Versorgung, Epidemiologie und Forschung bedeutet, seien zunächst die Ziele der epidemiologischen und klinischen Registrierung skizziert. Die epidemiologische Registrierung zählt das Auftreten der unterschiedlichen Krebserkrankungen und ihrer Stadien in einer Bezugsbevölkerung. Die systematische Erfassung der Leichenschaucheine und der Abgleich mit den Melderegistern erlaubt zudem die Berechnung von Überlebenszeiten und über die Death Certificate Notification (DCN) das Aufdecken von Meldelücken sowie die Berechnung des Anteils von Death Certificate Only (DCO), der einer der Qualitätsparameter der epidemiologischen Krebsregistrierung ist.

Dementsprechend sind aus epidemiologischen Registern Aussagen über zeitliche Trends und regionale Häufungen der Morbidität und Mortalität möglich. Die Interpretation solcher Zahlen kann zu Hypothesen über regionale oder allgemeine Risiken führen, die im Rahmen weiterführender Analysen und Studien erhärtet oder widerlegt werden. Epidemiologische Register sind Partner für Kohorten, in denen Risiken für Krebsentwicklung und -entdeckung erforscht werden. Sie können Aussagen über die Effizienz von Früherkennungs- und Präventionsmaßnahmen machen.

Die klinische Registrierung ist hingegen wesentlich detaillierter, da außer der Kategorisierung der Erkrankung auch Therapien, Therapieerfolge und -folgen sowie Statusangaben über den kompletten Verlauf der Erkrankung erfasst werden. Damit sind Aussagen zur Struktur-, Prozess- und Ergebnisqualität der onkologischen Behandlung möglich. Klinische Register sind wichtige Partner der zertifizierten Zentren in der Onkologie. Sie sind Ausgangspunkt zahlreicher klinischer Fragestellungen, die teilweise aus den Daten selbst, teilweise auch mit Hilfe von Zusatzdaten beantwortet werden. Auf Grund der Datentiefe leisten sie auch einen Beitrag zur individuellen Versorgung von Patienten, denn sie stellen bei Bedarf Übersichten zum Patienten zusammen, helfen bei der Organisation der Nachsorge oder unterstützen die Durchführung von Tumorkonferenzen.

Es gilt also für beide Registertypen, dass es Grenzen der aus eigenen Daten beantwortbaren Fragestellungen gibt und es dem Registerprinzip entspricht, dass ein deutlicher Mehrwert in der Kooperation mit Studien und/oder anderen Kohorten besteht.

Das KFRG ist als § 65c des SGB V realisiert. Darin ist festgelegt, dass die Krankenkassen mit einem maximalen Anteil von 90 Prozent die Arbeit der klinischen Krebsregister fördern. Die Krankenkassen haben dafür Förderrichtlinien mit Mindestanforderungen an die Krebsregister festgelegt, die erstmalig im Jahr 2018 überprüft werden. Die Deutsche Krebshilfe fördert ebenfalls mit Millionenbeträgen den Auf- und Ausbau der neuen Strukturen.

Es war relativ schnell klar, dass der Aufbau der klinischen Registrierung unter Nutzung der bestehenden Strukturen erfolgen sollte, um zum einen den Bestand der epidemiologischen Registrierung nicht zu gefährden und die bestehenden Meldewege zu nutzen, zum anderen um die vorhandenen Kompetenzen zu erhalten. Auch der enge Zeitrahmen bis 2018 erlaubt letztlich keine komplette Neustrukturierung.

Dies führt in der Praxis dazu, dass in den genannten „westlichen Ländern“ die epidemiologischen Register zusätzlich die klinische Registrierung übernehmen, während im Osten die klinischen Register weiterhin die Basis der Erfassung bilden und wie bisher die epidemiologischen Daten an das Landeskrebsregister (Bayern) bzw. das Gemeinsame Krebsregister der neuen Bundesländer und Berlins weiterleiten. Berlin, das bisher keine flächendeckende klinische Registrierung hatte, wird dabei mit Brandenburg zusammen ein Register führen, welches die zahlreichen überlappenden Versorgungswege optimal abbildet.

## **Anforderungen**

Die Tatsache, dass die Versorgungswege von Patienten Länder- und damit auch Registergrenzen überschreiten, führt zu einer wesentlichen Anforderung an die neuen Registerstrukturen. Für die Leistungserbringer ist es wichtig, dass beispielsweise zur Darstellung der Ergebnisqualität der Verlauf der Erkrankung nachvollzogen werden kann, und zwar unabhängig davon, wo der Patient weiterhin betreut wird. Für bevölkerungsbezogene Fragestellungen müssen hingegen die Daten am Wohnortregister zusammenfließen, auch wenn der Patient in einem anderen Registerbereich behandelt wurde. Das heißt, dass der Datenaustausch zwischen Registern reibungslos und zeitnah erfolgen muss. Die gesetzlichen und inhaltlichen Voraussetzungen dafür sind mit der voraussichtlichen Einführung der Meldepflicht in allen Bundesländern und der einheitlichen Definitionen der Inhalte (ADT-GEKID-Basisdatensatz und seine XML-Umsetzung) gegeben, wenn auch die technische Komplexität der Verarbeitung unter Umständen hoch ist.<sup>12</sup> Von verschiedenen Seiten wird eine stärkere Orientierung der Datenübertragung von und zu den Registern an internationalen und etablierten Standards wie etwa HL7 und IHE

---

<sup>12</sup> Siehe <http://www.tumorzentren.de/onkol-basisdatensatz.html>.

gewünscht. In der TMF wurde der hier entstandene Diskussionsbedarf im Herbst 2016 in zwei Workshops aufgegriffen.

### *Ausgabefunktionen von Registern*

Der § 65c des SGB V, der die Anforderungen an die Krebsregister beschreibt, legt unter anderem fest:

- ▶ die Rückmeldung der Auswertungsergebnisse an die einzelnen Leistungserbringer,
- ▶ die Förderung der interdisziplinären, direkt patientenbezogenen Zusammenarbeit bei der Krebsbehandlung,
- ▶ die Beteiligung an der einrichtungs- und sektorenübergreifenden Qualitätssicherung des Gemeinsamen Bundesausschusses [...],
- ▶ die Zusammenarbeit mit Zentren in der Onkologie sowie
- ▶ die Bereitstellung notwendiger Daten zur Herstellung von Versorgungstransparenz und zu Zwecken der Versorgungsforschung.

Das bedeutet, dass mit den Daten intensiv gearbeitet werden soll. Neben diversen Auswertungsfunktionen lässt sich ausdrücklich die Bereitstellung von Daten zu einem einzelnen Patienten ableiten. Die konkrete Ausgestaltung der Register wird allerdings den Ländern und deren Gesetzgebung überlassen.

So ist zu feststellen, dass die Organisationsform von Vertrauensbereich, Registerbereich und Auswertungsstelle in den bereits jetzt verabschiedeten Gesetzen von Land zu Land variiert. Auch die Regularien für die Kooperation mit Einrichtungen außerhalb der Register zur Beantwortung wissenschaftlicher Fragestellungen (z. B. eine möglicher Herausgabe bestimmter Daten) sind nicht einheitlich geregelt. Weitere Diskussionsfelder sind das Maß der künftig möglichen, versorgungsunterstützenden Funktionen und die Art der Erfassung. Im Westen liegen relativ zentralistische Konzepte vor, während versorgungsunterstützende Funktionen eher eine räumliche Nähe erfordern.

In den bestehenden klinischen Registern ist andererseits teilweise fraglich, ob noch alle unterstützenden Funktionen im Rahmen der Fallpauschale finanzierbar sind. Außerdem zeigt sich eine Tendenz zu einer starken Abschottung der Register aus Datenschutzgründen. Dies führt dazu, dass Register nicht mehr kooperativ geführt werden können. Bisher konnten speziell geschulte Dokumentationskräfte direkt an einem Datenbestand arbeiten und so Daten der Patienten ihrer Einrichtung unmittelbar für die Durchführung von Tumorkonferenzen und die Darstellung von synoptischen Krankengeschichten nutzen. Das wird zunehmend verboten, mit der Wirkung, dass kooperierende Einrich-

tungen eigene Dokumentationen aufbauen müssen, aus denen dann heraus gemeldet wird. Daten aus dem Register können nur über ein komplexes Verfahren abgerufen werden. Das ist nicht nur volkswirtschaftlich ineffektiv, sondern gefährdet auch eines der wesentlichen vom Gesetzgeber im SGB V formulierten Ziele.

Kritisch wird weiterhin diskutiert, welche Auswirkungen die Forderung der elektronischen Datenübermittlung vom Leistungserbringer (Melder) an das Krebsregister in den Förderrichtlinien des GKV-Spitzenverbandes hat. Bisher konnte in den bestehenden klinischen Registern die Codierung auf der Basis von Dokumenten wie Arztbriefen in den Registern vorgenommen werden, was die meist problematische Codierung durch Ärzte vermeidet. Durch Plausibilitätsprüfungen bereits strukturierter Daten kann aber nur die inhaltliche Widerspruchsfreiheit, nicht jedoch die Richtigkeit von Meldungen geprüft werden.

#### *Die § 65c-Plattform der Register*

Um die durch heterogene Gesetzgebung bewirkten Folgen abzumildern und wenigstens auf operativer Ebene zu weitgehend einheitlichen Verfahren zu kommen, wurde eine Plattform gegründet, auf der sich die KFRG-Register austauschen und z. B. Empfehlungen für Regelungen zur Vergütbarkeit von Meldungen oder den Austausch von Daten zwischen den Registern sowie Hilfen für die Softwareindustrie zur Umsetzung von elektronischen Meldungen erarbeiten.

So zeigte es sich beispielsweise, dass die Softwareindustrie Schwierigkeiten hatte, allein auf der Basis des ADT-GEKID-XML-Datensatzes und seiner Beschreibungen Lösungen zu implementieren. In einem Umsetzungsleitfaden wurden die Verständnisprobleme gezielt aufgearbeitet und Lücken in der Beschreibung behoben. Außerdem ist zu beobachten, dass über den Umsetzungsleitfaden hinaus Entwicklungen begonnen werden, den ADT-GEKID-Datensatz in bestehende Standardbeschreibungen – zum Beispiel auf der Basis von HL7 und IHE – zu überführen, was wiederum die Umsetzbarkeit erleichtern soll.

### **Datenqualität**

Bei der Betrachtung der Datenqualität spielen folgende Aspekte eine wesentliche Rolle [5]:

- ▶ Vollzähligkeit: die vollzählige Erfassung der relevanten Erkrankungen und der dazu zu dokumentierenden Ereignisse
- ▶ Vollständigkeit: das Maß dafür, inwieweit die Ereignisse adäquat dokumentiert sind
- ▶ Richtigkeit: Übereinstimmung der Dokumentation mit der Realität

Was die Vollzähligkeit der Erkrankungen betrifft, existieren in der epidemiologischen Registrierung etablierte Methoden: die systematische Einbindung aller Pathologen, die Nachrecherche nur über Leichenschauschein bekannt gewordene Erkrankungen (DCN), das Ausweisen des nach Nachrecherche verbliebenen DCO-Anteils und den Mortalitäts-/Inzidenz-Index, mit dessen Hilfe abgeschätzt werden kann, bei welcher Mortalität wie viele klinische Ereignisse zu erwarten gewesen wären. Bei der klinischen Registrierung kommt es jedoch zusätzlich darauf an, auch alle Therapien und Folgeereignisse zu erfassen. Die Methoden zur Qualitätssicherung dieser Erfassung sind noch nicht systematisch beschrieben. Ein mögliches Hilfsmittel wäre aber die Berücksichtigung von Routinecodierungen, die in einem wenigstens wahrscheinlichen Kontext mit Krebserkrankungen stehen können: ICD-Codes, OPS-Codes in zeitlichem Zusammenhang mit ICD-Codes sowie die Leistungserfassung in onkologiespezifischen Leistungsstellen (Strahlentherapie, onkologische Ambulanzen, Chemotherapiezubereitung).

Zur Erhöhung der Vollständigkeit hilft der Einsatz von Pflichtfeldern nur bedingt weiter. In Abhängigkeit von der Art der Erkrankung sind nicht immer alle abzufragenden Inhalte relevant. Beispielsweise ist die Dokumentation der TNM-Klassifikation bei systemischen Erkrankungen nicht sinnvoll, bei Revisionsoperationen kann keine R-Klassifikation angegeben werden und es verbieten sich in bestimmten Situationen aufwändige Untersuchungen, deren Ergebnisse anderenfalls festzuhalten wären. Diese Tatsache erschwert die Definition von immer wieder geforderten automatisierten Prüfungen in Erfassungssystemen. Auf der anderen Seite kann es die Dokumentation extrem vereinfachen, wenn in einem bestimmten Umfeld nur ausgewählte Situationen auftreten.

Die optimale Überprüfung der Richtigkeit würde einen wenigstens stichprobenhaften Abgleich der Registerdaten mit den Krankenakten erfordern. Das ist aber aus Ressourcengründen kaum erreichbar. Manchmal kommen Informationen zum gleichen Sachverhalt aus unterschiedlichen Quellen. Dann kann die inhaltliche Stimmigkeit überprüft werden. Dies ist insbesondere für den diagnostischen Prozess zu erwarten, da an der kompletten Diagnose eine Reihe von Leistungserbringern (Meldern) beteiligt ist.

### **Elektronische Meldung**

Wie bereits erwähnt, soll die Meldung bevorzugt elektronisch erfolgen. Mit der Einführung flächendeckender klinischer Krebsregister ist nicht nur der zu meldende Datensatz inhaltlich gegenüber der epidemiologischen Registrierung stark erweitert worden, sondern auch die Häufigkeit der Meldungen. Letztlich ist es ja das Ziel, den kompletten Verlauf einer Krebserkrankung abzubilden.

Abbildung 4.5 zeigt auf der Zeitachse die unterschiedlichen Phasen einer Erkrankung. Im Rahmen der Primärdiagnostik sind unterschiedliche Leistungserbringer (Ärzte/Melder) daran beteiligt, den Ausgangsstatus festzustellen und zu beschreiben. Da kein Beteiligter die Kontrolle über das Meldeverhalten der vorangehenden und nachfolgenden Beteiligten besitzt, ist auch jeder, der etwas zur Diagnostik beiträgt, grundsätzlich verpflichtet, den aktuellen Stand der Diagnostik zusammenzufassen und zu melden.

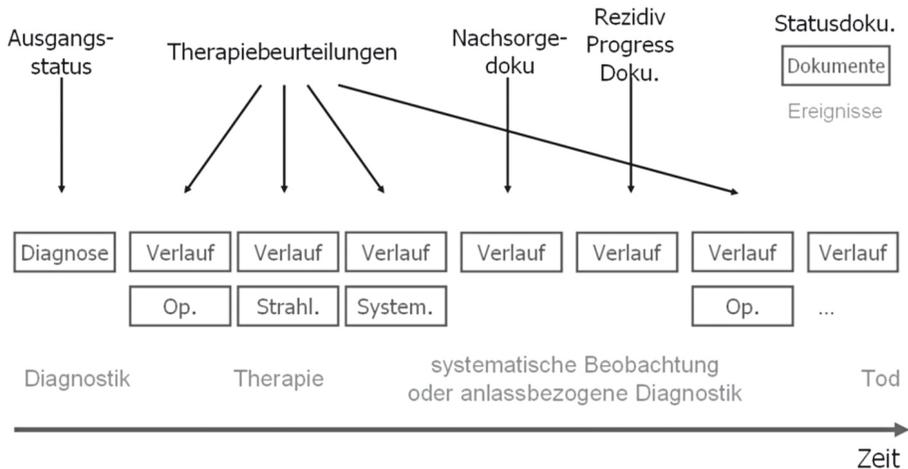


Abbildung 4.5: Ereignisse im Verlauf einer Krebserkrankung und deren Abbildung auf Meldeinhalte.

Je nach Erkrankungstyp und Ausbreitung erfolgt in aller Regel eine Therapie, die wiederum bei unterschiedlichen Leistungserbringern stattfinden kann. Anschließend wird der Patient entsprechend den Empfehlungen – z. B. von Leitlinien – mehr oder weniger aktiv nachbeobachtet. Dabei kann im Rahmen der Nachsorge eine fortdauernde Tumorfreiheit festgestellt werden. Es kann aber auch sein, dass der Patient ein Fortschreiten der Erkrankung oder ein Rezidiv erleidet und erneut behandelt wird. Unabhängig davon, ob der Tod durch die Tumorerkrankung verursacht wurde oder nicht, wird er mit den Todesursachen registriert.

Über der Zeitachse sind die entsprechenden Meldedokumente dargestellt. Während die untere Zeile die therapie relevanten Dokumente zeigt, sind in der oberen Zeile neben der initialen Statusbeschreibung die Verlaufs dokumente zu sehen, die zu unterschiedlichen Anlässen erzeugt werden. Hier sind in jedem Fall Statusänderungen von Interesse, also der Eintritt von Tumorfreiheit oder Rezidiv/Progression/Transformation. Aus methodischen Gründen soll aber in einem Teil der Bundesländer auch eine Statusfortschreibung gemeldet werden.

Grundsätzlich ist jeder Leistungserbringer verpflichtet, die bei ihm entstandenen Daten zu melden. Da ein Patient gleichzeitig oder nacheinander an mehreren Tumorerkrankungen leiden kann, ist jeweils eine den Tumor eindeutig identifizierende Grundinformation mit zu melden. Um die Daten auch für aktuelle Aufgabenstellungen verwenden zu können, ist eine zeitnahe Meldung erforderlich.

Bei der elektronischen Meldung ist anzustreben, dass vorhandene Routinedaten weitgehend genutzt werden, da eine Mehrfachdokumentation im Allgemeinen akzeptanzschädigend ist. In der Praxis kommen folgende Softwarekonzeptionen zum Einsatz.

#### *Volle Integration der Datenerfassung in das Arbeitsplatzsystem*

Neben der Nutzung des vertrauten Benutzerinterfaces ist der Vorteil dieser Konzeption, dass es in den Informationssystemen bereits jetzt eine Reihe möglicher Auslöser (Trigger) für die Meldung gibt, die die Vollzähligkeit unterstützen:

- ▶ Erfassung in onkologiespezifischen Leistungsstellen (bestimmte Abteilungen wie Strahlentherapie, spezielle Tumorambulanzen und onkologische Praxen)
- ▶ Onkologiespezifische Leistungsdokumentation (ICD-Codierung, OPS-Codes mit hoher Tumorspezifität bzw. im Kontext einer Tumor-ICD, Bestrahlungsplanungs- und -verifikationssysteme, Systeme zur Chemotherapieplanung, spezielle Leistungsziffern oder OPS-Codes für Chemo- oder Strahlentherapien)
- ▶ Befundberichte aus der Pathologie mit vordefinierten kritischen Angaben (ICD-O-Codes, TNM)

Allerdings ist, sofern diese Trigger genutzt werden, der Umfang der vorhandenen Informationen und ihr Strukturierungsgrad im Regelfall nicht ausreichend, so dass zusätzliche strukturierte Informationen erfragt werden müssen. Sofern Codierungen vorgenommen werden müssen, zeigt die Erfahrung, dass die ärztliche Codierqualität derjenigen von speziellen Dokumentationskräften meist unterlegen ist. In der Praxis werden häufig in den Krankenhausinformationssystemen (KIS) und Arztpraxisinformationssystemen (AIS) explizite Zusatzformulare für die Tumordokumentation hinterlegt und so gut wie möglich aus bestehenden Daten vorausgefüllt. Diese Lösung kann zudem dazu führen, dass die Menge strukturiert erfasster Informationen in den Systemen zunimmt, was sich wiederum positiv auf deren Auswertungsmöglichkeiten auswirkt.

#### *Anbinden von Zusatzmodulen*

Für die externe Qualitätssicherung werden häufig spezielle Programme eingesetzt, die

ebenfalls die Daten des vorhandenen Informationssystems auf Triggercodes filtern können, um dann in vorab festgelegten Fällen Dokumentationsformulare mit vorbelegten Feldern anzubieten. Der Dokumentationsprozess für die externe Qualitätssicherung findet allerdings häufig nicht zeitnah statt, da die Daten lediglich in größeren Intervallen abgefragt werden. Dennoch wird auch dieser Mechanismus stellenweise eingesetzt, um Meldungen an Krebsregister zu generieren, da sich diese Programme in hohem Grad durch externe Spezifikationsdateien steuern und so an den Anwendungsfall der elektronischen Meldung an Krebsregister anpassen lassen.<sup>13</sup> Manche Register bieten auch eigene Programme an, an die aus KIS oder AIS in vergleichbarer Form Daten übergeben werden können. Diesen Lösungen ist gemeinsam, dass sie ihr eigenes Benutzerinterface haben, in dem die erforderlichen Ergänzungen vorgenommen werden müssen. Bezüglich Dokumentationsqualität gilt das bereits im vorherigen Abschnitt gesagte.

### *Eigenständige Tumordokumentationssysteme*

In diesem Fall handelt es sich um komplett eigenständige Erfassungssysteme, die aber meist Schnittstellen zu KIS anbieten, damit wenigstens einige der bereits andernorts erfassten Daten übernommen werden können. In aller Regel werden diese durch spezielle Dokumentationskräfte bedient. Wegen der inhaltlichen Nähe zu den allgemeinen Regeln der Tumordokumentation und bei entsprechender Qualifikation dieser Kräfte liefern diese Systeme im Allgemeinen eine gute Datenqualität.

## **Bewertung und Handlungsbedarf**

In der Erhebung von 2012 zu Registern bei den Mitgliedsverbänden der TMF lag der Betrieb von IT-Komponenten vorwiegend in der Hand universitärer oder Universitäts-naher Einrichtungen. Dennoch wurde häufiger als bei den vorangehenden Erhebungen auf die Möglichkeit des Outsourcings zurückgegriffen. Bei rund der Hälfte der beschriebenen Betriebskonzepte wurde zumindest teilweise diese Möglichkeit genutzt. Es dominierte dabei die Auslagerung von zentralen Hardwarekomponenten. Bei den Softwareprodukten waren Eigenentwicklungen unter Nutzung freier Software führend. Unter den kommerziellen Produkten fanden sich vorwiegend EDC-Systeme aus dem Umfeld klinischer Studien, die in einem zum Teil langjährigen Prozess von den Anwendern an die Spezifika von Registern angepasst wurden. Die Planung von Registern fand weitgehend ohne spezifische IT-Unterstützung statt.

---

13 Das AQUA-Institut stellt hierfür auch eine Spezifikation zur Verfügung, die von Herstellern von Systemen zur externen Qualitätssicherung für diesen Anwendungsfall genutzt werden kann, siehe <https://www.aqua-institut.de/de/projekte/spezifikation-krebsregister/>.

Die aktuelle Zusammenstellung von spezifischen Softwareprodukten für Register und Kohorten zeigt, dass inzwischen ein Markt mit kommerziellen, aber auch mit frei verfügbaren Angeboten entstanden ist. Aufgabe der TMF wird es sein, diese Zusammenstellung kontinuierlich zu ergänzen und zu aktualisieren. Vergleichbares ist auch für den Bereich der Tumordokumentation und für Krebsregister wünschenswert. Zur Beschreibung von Softwareprodukten, zur Begleitung einer Produktauswahl sowie zur Unterstützung des IT-Managements steht mit dem dargestellten Anforderungskatalog zur IT-Unterstützung von Kohorten und Registern ein passendes Instrument zur Verfügung.

Zunehmende Aufmerksamkeit findet das Thema Datenqualität [22]. Für Register, Kohorten und Data Repositories liegt eine umfangreiche Empfehlung mit der Beschreibung von 51 Indikatoren zur Qualität von Daten vor [5]. Nicht bekannt sind allerdings Umfang, Art und Marktlage einer IT-Unterstützung dieser Aufgabe. Auch hier besteht Bedarf an einer systematischen Aufarbeitung.

Register sind langfristig angelegte Vorhaben mit prinzipiell unbegrenzter Laufzeit und Nachverfolgung der eingeschlossenen Personen bis zu deren Versterben. Die Zahl der Teilnehmer ist zumeist vierstellig. Der Wunsch nach möglichst vielen Informationen über die Teilnehmer zeigt sich in dreistelligen Zahlen von zu erfassenden Merkmalen. Gemessen an diesen Größenordnungen muss die 2012 genannte Zahl von im Median nur einer Vollkraftstelle für den zentralen Betrieb eines Registers als Unterfinanzierung bewertet werden. Verlässliche oder zumindest begründete Schätzungen für eine sinnvolle finanzielle Ausstattung liegen allerdings nicht vor. Für weitere Fördermaßnahmen<sup>14</sup> oder verlässliche Betriebs- und Geschäftsmodelle wären solche Zahlen gleichwohl von zentraler Bedeutung.

Bei den erreichten Ergebnissen und den angestrebten Zielen zeigen sich Verschiebungen. In einer Erhebung der TMF von 2009 dominierte noch die Unterstützung klinischer Studien durch eine Hypothesengenerierung, eine Rekrutierungsunterstützung und eine Kopplung des Registers mit klinischen Studien. Die Ermittlung epidemiologischer Kennzahlen und die Unterstützung der Qualitätssicherung – also klassische eigenständige Ziele von Registern – wurden 2012 erstmals gleich häufig als Ergebnisse der befragten Register genannt. Nach einem Aufbau vieler Register im Rahmen der Förderung von Kompetenznetzen in der Medizin hatte sich seit 2008 in einer zweiten Welle eine Vielzahl weiterer Register etabliert. Dies belegt, dass Register in der deutschen Forschungslandschaft als ein wichtiges Werkzeug zur Unterstützung von patientenorientierter Forschung und Gesundheitsversorgung angesehen werden. Per-

---

<sup>14</sup> Siehe etwa die aktuelle Ausschreibung zu Registern der Versorgungsforschung unter <http://www.gesundheitsforschung-bmbf.de/de/6663.php>.

spektivisch deutet sich mit der Einbettung klinischer Studien in Register ein weiterer Bedeutungszuwachs an [vgl. 23].

Angesichts der zunehmenden Bedeutung von Registern und registergestützter Forschung stellt sich zunehmend auch die Frage nach ihrer IT-Integration in die Strukturen der Versorgung und Forschung. Dies ist nicht nur, aber in aktueller und besonderer Weise auch ein Thema für die aktuell entstehenden klinischen Krebsregister. Hier wird aktuell von verschiedenen Seiten diskutiert, wie eine Datenübertragung zwischen Registern und zwischen Registern und anderen Datenbeständen technisch gestaltet werden kann. Hierbei wird eine Orientierung an etablierten Standards wie etwa HL7 und IHE gewünscht. In der TMF wurde der entstandene Diskussionsbedarf im Herbst 2016 in zwei Workshops aufgegriffen. Gerade die Einbettung dieses Themas in die Arbeit der TMF spricht dafür, dass von den hier dringend notwendigen Fortschritten am Ende nicht nur die klinischen Krebsregister, sondern auch andere Register profitieren können.

Aus der Initiative von DNVF und TMF zum Aufbau eines Registerportals liegen erste Ergebnisse vor [13]. In einem Konsensverfahren unter Einbindung von Expertinnen und Experten beider Gesellschaften wurde ein Kerndatensatz zur Beschreibung von Registern und Kohorten abgestimmt. Dieser umfasst allgemeine Angaben zum Vorhaben wie Name und Akronym, eine freitextliche Kurzbeschreibung, Informationen zur Struktur des Registers oder der Kohorte wie die geplante Laufzeit und die Region der Rekrutierung sowie Hinweise auf Dienstleistungen und entsprechende Regelungen. Der ausführliche Merkmalskatalog und das Datenschutzkonzept können optional zum Download angeboten werden. Die entworfenen Verfahrensregeln gehen davon aus, dass die Nutzung des Registerportals für Interessenten wie für Betreiber medizinischer Register und Kohorten gebührenfrei ist. Die Verantwortung für die laufende Aktualisierung der Beschreibung der Register und Kohorten würde weitestgehend an die für die registrierten Einrichtungen oder Projekte jeweils Verantwortlichen delegiert. Das Datenmanagement des Registerportals selbst wäre dann für formale Prüfungen und ggf. auch Korrekturen verantwortlich. Mit der Festlegung der Metadaten und der Beschreibung von Verfahrensregeln sind die konzeptionellen Voraussetzungen für die Inbetriebnahme des Registerportals ausformuliert. Notwendig ist jetzt die Finanzierung von Aufbau und Betrieb eines solchen Übersichtsregisters. Nicht zuletzt ist eine solche Datenquelle auch für die weitere Verfolgung des IT-Bedarfs von Registern und Kohorten unverzichtbar.

Während private Stellen und kommerziell orientierte Akteure seit längerem Interesse an bestehenden Datensammlungen im Gesundheitswesen zeigen, häufig mit dem Stichwort „Big Data“ in Verbindung gebracht, wird jetzt auch seitens öffentlicher Einrich-

tungen die Bedeutung einer Übersicht zu vorhandenen und z. T. auch mit öffentlichen Mitteln geförderten Datenbeständen artikuliert. Als Beleg hierfür kann zum einen ein vom Büro für Technikfolgenabschätzung beim Deutschen Bundestag 2014 vergebenes Gutachten zum Data-Mining in der Medizin dienen, welches explizit eine umfassende Darstellung der relevanten Datenbestände und ihrer Zugänglichkeit zum Inhalt haben sollte. Zum anderen fördert das BMBF ein Begleitprojekt zur Fördermaßnahme „Medizininformatik“, dessen Auftrag auch die Erstellung einer „Datenlandkarte“ beinhaltet.<sup>15</sup> Ein Übersichtsregister mit einer laufend aktualisierten Beschreibung vorhandener Register und Kohorten in Deutschland wäre eine zentrale Voraussetzung dafür, auf diesen Informationsbedarf eine nachhaltige Antwort zu finden.

## Danksagung

Das Projekt „Von der Evaluierung zur Konsolidierung: Anforderungen an Kohortenstudien & Register IT (KoRegIT)“ wurde von der TMF mit der Nummer V078-01 unter Projektleitung von Frau Claudia Michalik, Köln, gefördert. Für die Informationen zu OpenClinica und REDCap danken wir Herrn Matthias Löbe, Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE), Universität Leipzig. Ebenso danken wir den Unternehmen für die Informationen zu der von ihnen angebotenen Software.

## Literatur

1. Glicklich, R.E., Dreyer, N.A., Leavy, M.B., Hrsg. *Registries for Evaluating Patients Outcomes: A User's guide*. 3. Aufl. 2014, Agency for Healthcare and Research Quality, Rockville, MD.
2. Neugebauer, E.A.M., Stausberg, J., *Was Register leisten können und was nicht. Sicht der AG Register des Deutschen Netzwerks Versorgungsforschung e. V. (DNVF)*. Der Unfallchirurg, 2016. **119**(6): S. 493-500.
3. Müller, D., Augustin, M., Banik, N. et al., *Memorandum Register für die Versorgungsforschung*. Gesundheitswesen, 2010. **72**(11): S. 824 - 839.
4. DNEbM *Glossar zur Evidenzbasierten Medizin*. 2011. Deutsches Netzwerk Evidenzbasierte Medizin e.V., <http://www.ebm-netzwerk.de/was-ist-ebm/grundbegriffe/glossar> (Abruf: 2016-11-20).
5. Nonnemacher, M., Stausberg, J., Nasseh, D., *Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern*. 2. Aufl. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.

---

<sup>15</sup> Das Begleitprojekt wurde an die TMF vergeben, die dies in enger Kooperation mit MFT und VUD bearbeitet.

6. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
7. Schwanke, J., Rienhoff, O., Schulze, T.G., Nussbeck, S.Y., *Suitability of Customer Relationship Management Systems for the Management of Study Participants in Biomedical Research*. *Methods of Information in Medicine*, 2013. **52**(4 2013): S. 340 - 350.
8. Muscholl, M., Lablans, M., Hirche, T., Wagner, T., Ückert, F., *OSSE - Open-Source-Registersystem für Seltene Erkrankungen in der EU*, in *Big Data und Forschungsinfrastruktur - Perspektiven für die Medizin. Abstractband GMDS 2014*, Hrsg.: S.Y. Nußbeck, S. Wolff, and O. Rienhoff. 2014, Schattauer, Stuttgart. S. 317-319.
9. Altmann, U., Dudeck, J., *The Giessen Tumor Documentation System (GTDS)--review and perspectives*. *Methods Inf Med*, 2006. **45**(1): S. 108-115.
10. Michalik, C., Dress, J., Ngouongo, S. et al., *Requirements and tasks of cohorts and registers, the German KoRegIT project*. *Stud Health Technol Inform*, 2014. **205**: S. 1085-1089.
11. Niederländer, C., Wahlster, P., Kriza, C., Kolominsky-Rabas, P., *Registries of implantable medical devices in Europe*. *Health Policy*, 2013. **113**(1-2): S. 20-37.
12. NAKO *The National Cohort. A prospective epidemiologic study resource for health and disease research in Germany*. 2011. Nationale Kohorte e.V., <http://nako.de/wp-content/uploads/2015/07/Wissenschaftliches-Konzept-der-NAKO2.pdf> (Abruf: 2016-11-20).
13. Stausberg, J., Semler, S.C., Neugebauer, E.A.M., *Ein Register für Register und Kohorten: Empfehlungen zu Metadaten und Verfahrensregeln*. *Das Gesundheitswesen*, 2014. **76**: S. 865-873.
14. Stausberg, J., Schütt, A. *Register bei den Mitgliedsverbänden der Telematikplattform für medizinische Forschungsnetze (TMF e.V.) Ergebnisse einer Umfrage im Dezember 2007*. 2008. TMF e.V.
15. Stausberg, J., Altmann, U., Antony, G., Drepper, J., Sax, U., Schütt, A., *Registers for Networked Medical Research in Germany: Situation and prospects*. *Appl Clin Inform*, 2010. **1**(4): S. 408-418.
16. Korfkamp, D., Gudenkauf, S., Rohde, M. et al., *Opening up Data Analysis for Medical Health Services: Cancer Survival Analysis with CARESS*, in *Transactions on Large-Scale Data- and Knowledge-Centered Systems XXVI. Special Issue on Data Warehousing and Knowledge Discovery*, Hrsg.: A. Hameurlain, et al. 2014, Springer, Berlin Heidelberg. S. 89-107.

17. Bialke, M., Bahls, T., Havemann, C., Piegsa, J., Weitmann, K., Wegner, T., Hoffmann, W., *MOSAIC – A Modular Approach to Data Management in Epidemiological Studies*. *Methods of Information in Medicine*, 2015. **54**(4): S. 364-371.
18. Stuke, K., Flachenecker, P., Zettl, U.K. et al., *Symptomatology of MS: results from the German MS Registry*. *J Neurol*, 2009. **256**(11): S. 1932-1935.
19. Antony, G., Buckow, K., Khil, L., Stroet, A. *Entwicklung eines gemeinsamen modularen Datensatzes für die Studien des KKNMS und der DMSG*. 55. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS). 2010. Mannheim.
20. Stahmann, A., Meyer, C., Lee, M., Buckow, K., *PRO-Dokumentation von Lebensqualitätsdaten im MS-Register der Deutschen Multiplen Sklerose Gesellschaft (DMSG)*, Bundesverband e.V. mdi, 2015. **2015**(17): S. 18-19.
21. Flachenecker, P., Buckow, K., Pugliatti, M. et al., *Multiple sclerosis registries in Europe - results of a systematic survey*. *Mult Scler*, 2014. **20**(11): S. 1523-1532.
22. Malin, J.L., Keating, N.L., *The cost-quality trade-off: need for data quality standards for studies that impact clinical practice and health policy*. *J Clin Oncol*, 2005. **23**(21): S. 4581-4584.
23. Frobert, O., Lagerqvist, B., Gudnason, T., Thuesen, L., Svensson, R., Olivecrona, G.K., James, S.K., *Thrombus Aspiration in ST-Elevation myocardial infarction in Scandinavia (TASTE trial). A multicenter, prospective, randomized, controlled clinical registry trial based on the Swedish angiography and angioplasty registry (SCAAR) platform. Study design and rationale*. *Am Heart J*, 2010. **160**(6): S. 1042-1048.

# Erhebung, Management und Verarbeitung digitaler Bilder

Thomas Franke, Ulrich Sax

Mitarbeit: Johannes Bernarding, Peter Hufnagl, Hans-Peter Mainzer, Bartłomiej Marzec, Peter Mildenberger

## Anwendungsbereich

Die Bildgebung und die softwaregestützte Bildverarbeitung sind in der Medizin seit vielen Jahren fest etabliert. Bei vielen Indikationen ist die Bildgebung aus der modernen Diagnostik nicht mehr wegzudenken. In der Forschung werden Bildgebung und Bildverarbeitung sowohl zur Bestimmung von Surrogat-Endpunkten, als auch zur Analyse von Krankheitsmechanismen eingesetzt. In diesem Kapitel werden die speziellen Herausforderungen, die Erhebung, Management und Verarbeitung medizinischer Bilddaten an die Informationstechnologie stellen, strukturiert und im Überblick dargestellt. Die Strukturierung folgt einem generischen Workflow der forschungsbezogenen Bildverarbeitung (siehe Abb. 5.1).

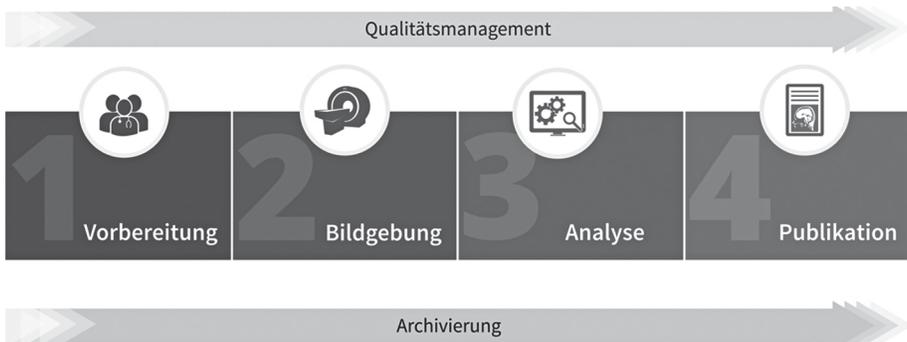


Abbildung 5.1: Der generische Workflow der forschungsbezogenen Bildverarbeitung, unterteilt in vier Hauptphasen und zwei phasenübergreifende Prozesse.

Der Prozess der medizinischen Bildverarbeitung in der Forschung kann grob in vier Hauptphasen unterteilt werden: 1) Vorbereitungsphase, 2) Bildgebungsphase, 3) Analysephase und 4) Publikationsphase. Während der Vorbereitungsphase geht es darum, die passende Bildgebungsmethode für die gegebene Fragestellung bzw. die gegebenen Zielp Parameter auszuwählen und die notwendigen Schritte zur Sicherung der gewünschten Qualität in die Wege zu leiten. Konkret umfasst diese Phase beispielsweise die

Auswahl der Modalität und der Messmethoden, die Erstellung von Messprotokollen und Standard Operating Procedures (SOPs) sowie – insbesondere bei multizentrischen Projekten – die Schulung des an der Bildgebung beteiligten Personals.

Die Bildgebung kann dann entweder im Behandlungszusammenhang, beispielsweise im Rahmen einer klinischen Studie oder exklusiv im Forschungskontext stattfinden. Im Anschluss an die Erhebung der Bilddaten erfolgt zunächst deren Speicherung. Je nach Projektdesign werden die Daten entweder lokal in der erhebenden Einrichtung oder, beispielsweise bei multizentrischen Studien mit zentraler Infrastruktur, bei einer anderen Institution gesammelt. Ist letzteres der Fall, so stellt die Datenübertragung eine Herausforderung dar, für die in Abhängigkeit von der jeweiligen Situation eine geeignete Lösung gewählt werden muss. In jedem Fall muss sichergestellt werden, dass die Daten vor der Übertragung adäquat pseudonymisiert und so weit wie möglich von identifizierenden Informationen bereinigt worden sind. Bei der Übertragung und der Pseudonymisierung von Bilddaten sind einige Besonderheiten zu beachten, die im Weiteren noch detailliert ausgeführt werden. Die eigentlichen Speicherlösungen variieren von der einfachen Dateisystemablage, über die Unterbringung im klinischen Picture Archiving and Communications System (PACS) bis hin zum Einsatz von spezialisierten Softwarelösungen für die Verwaltung von Bilddaten im Forschungskontext. Einige Vertreter der letztgenannten Variante werden am Ende dieses Kapitels kurz vorgestellt.

In der Analysephase werden die Bilddaten im Hinblick auf die ursprünglich definierten Zielkriterien ausgewertet. Teilweise werden die Bilder dazu zunächst softwaregestützt vorverarbeitet, bevor eine qualitative oder quantitative Auswertung stattfindet. Bei dieser Vorverarbeitung werden die Bilddaten mitunter in andere Dateiformate konvertiert, um Kompatibilität zu den Verarbeitungs- und Auswertungsprogrammen herzustellen. Darüber hinaus existieren für einige Zielparameter auch Algorithmen, die die eigentliche Messung automatisieren können. Im Anschluss an die Auswertung einzelner Datensätze erfolgt – oftmals in Kombination mit anderen Datentypen – die statistische Analyse der Ergebnisse. Die Vielzahl an möglichen Bilddaten und unterschiedlichsten Zielparametern hat im Bereich der Auswertungssoftware zur Entwicklung vieler Werkzeuge geführt, von denen eine Auswahl am Ende dieses Kapitels erwähnt wird.

Die letzte Phase des generischen Workflows umfasst die Veröffentlichung. Hierbei ist grundsätzlich zwischen der Publikation der Ergebnisse und der Veröffentlichung der Daten, beispielsweise im Sinne des Open-Access-Prinzips, zu unterscheiden. Letzteres wird bisher nur selten durchgeführt, gewinnt aber unter anderem aufgrund von Vorgaben der öffentlichen Förderer zunehmend an Bedeutung.

Im Idealfall werden diese vier Phasen von einem durchgehenden Qualitätsmanagementprozess, sowie einer durchgehenden Archivierung begleitet. Die Granularität der Archivierung kann situationsabhängig variieren. Grundsätzlich sollten jedoch nicht nur die erzeugten Bilddaten, sondern auch Dokumente, die die Begleitumstände beschreiben, beispielsweise Messprotokolle und SOPs archiviert werden. Wenn Bilddaten im Rahmen der Auswertung verarbeitet und dadurch verändert werden, bieten sich für die Archivierung zwei Optionen an: Entweder es werden alle Zwischenergebnisse aufbewahrt oder es wird sichergestellt, dass die Verarbeitung jederzeit reproduziert werden kann. Hierfür müssen eine Kopie der Softwareumgebung, in der die Verarbeitungen durchgeführt worden sind, sowie eine Beschreibung des Verarbeitungsprozesses archiviert werden.

## Kategorisierung der IT-Landschaft

Eine grundsätzliche Möglichkeit der Kategorisierung ergibt sich beim Blick auf den Einsatzzweck der verfügbaren Produkte: Es gibt Softwarelösungen für das Management der Daten, die in ihren grundlegenden Funktionen dem klinischen PACS ähneln, sowie Tools zur Verarbeitung der Bilder. Letztere können noch einmal unterteilt werden in Produkte, die ausschließlich zur Vorbereitung bzw. Vorverarbeitung der Daten genutzt werden und solche, die zur Visualisierung, Analyse und Auswertung dienen; teilweise decken Werkzeuge allerdings auch beide Bereiche ab. Hinzu kommen noch Programme, die in der Lehre oder zu Schulungszwecken eingesetzt werden.

Zusätzlich zum eigentlichen Einsatzzweck kann in der medizinischen Bildgebung auch zwischen der DICOM-Welt und der Non-DICOM-Welt unterschieden werden. Der Standard Digital Imaging and Communications in Medicine (DICOM<sup>1</sup>) ist ein wichtiger Eckpfeiler für die Interoperabilität von IT-Systemen im Gesundheitswesen. Der DICOM-Standard umfasst gleichzeitig Kommunikationsprotokolle, sowie die Spezifikationen für ein Dateiformat. Bildgebende Modalitäten, die an ein klinisches PACS angebunden sind, sind in der Regel selber DICOM-fähig oder werden über ein nachgeschaltetes Frame-Capture-Gerät DICOM-fähig gemacht. Zu den Bildgebungsverfahren, für die typischerweise DICOM-fähige Modalitäten eingesetzt werden, gehören unter anderem die Computertomographie (CT), die Magnetresonanztomographie (MRT), die Positronen-Emissions-Tomographie (PET) – inklusive der Hybridformen PET-CT und PET-MRT – sowie das digitale Röntgen und zunehmend auch die Ultraschall-Bildgebung. Auf der anderen Seite stehen Geräte, die überwiegend von Abteilungen eingesetzt werden, die nicht an ein PACS angebunden sind, sowie Geräte, die

---

1 <http://dicom.nema.org/>

ausschließlich in der präklinischen Forschung eingesetzt werden. Diese arbeiten häufig mit proprietären Dateiformaten und Kommunikationsprotokollen; beispielhaft seien an dieser Stelle Slide Scanner und digitale Mikroskope erwähnt, die überwiegend in der Pathologie zum Einsatz kommen.

Die Trennung zwischen diesen Welten prägt auch die IT-Landschaft. Über die Zeit wurde eine Vielzahl von Softwareprodukten – freie und kommerzielle – für das Management und für die Verarbeitung medizinischer Bilddaten entwickelt. Die meisten von ihnen sind jedoch entweder auf die Anwendung in der DICOM-Welt oder aber in der Non-DICOM-Welt spezialisiert; nur wenige überbrücken diesen Graben und werden in beiden Bereichen eingesetzt.

Bei der zugrundeliegenden Speicherarchitektur kann zwischen Kurzzeit- und Langzeitspeicher unterschieden werden. Es können auch ausgelagerte Speicherstrukturen bei externen Dienstleistern genutzt werden. Hierzu sind neben den einschlägigen Datenschutzbestimmungen insbesondere auch Datensicherheitsaspekte wie Verfügbarkeit, Schutz vor Datenverlust und Wiederherstellungszeiten zu beachten und über gesonderte Verträge sicherzustellen. Die Rentabilität einer solchen Auslagerung hängt stark von den Gegebenheiten des Einzelfalls ab.<sup>2</sup>

## Betreiber, Anwender und Nutzer

Üblicherweise verfügen Standorte von klinischen Forschungsprojekten über ein zentrales klinisches PACS. Dies wird in der Regel vom lokalen Rechenzentrum betrieben, manchmal aber auch von einem externen Dienstleister bereitgestellt. An dieses PACS sollten alle bildgebenden Modalitäten und bildbeziehenden Abteilungen der DICOM-Welt angeschlossen sein. Oftmals gibt es noch Ausnahmen bezüglich des Anschlusses einzelner Einrichtungen, wie beispielsweise einer Zahnklinik. Daneben gibt es im Bereich der Pathologie eine zunehmende Anzahl an bildgebenden Modalitäten außerhalb der DICOM-Welt, für die separate Speicher- und Managementlösungen betrieben werden.

Zentrale Archive für Forschungsbilddaten gibt es nicht an allen Standorten in Deutschland. Wenn ein solches System vorhanden ist, kann es entweder vor Ort betrieben – beispielsweise durch eine Core Facility – oder von einem externen Dienstleister zur Verfügung gestellt werden.

---

2 Siehe hierzu auch das Kapitel „Cloud-Computing in der medizinischen Forschung“ in diesem Buch.

In der klinischen Forschung sind die primären Anwender eines Archivs für Forschungsbilddaten Abteilungen, in denen Bilddaten zur Durchführung von klinischen Studien erhoben bzw. verarbeitet werden, sowie Forschungsverbände mit entsprechenden Projekten. Die Nutzer sind in erster Linie Ärzte, Technische Assistenten und Study Nurses, die die Daten erheben und in das System einstellen, sowie Forscher, die anhand der etablierten Datenbasis eigene Auswertungen zur Beantwortung ihrer Forschungsfragen durchführen. Im Kontext von standortübergreifenden, multizentrischen Forschungsprojekten oder Kooperationen sind sowohl Anwender als auch Nutzer oftmals externe Personen bzw. Organisationen, die nicht zu der Einrichtung gehören, die das System betreibt bzw. die Ressourcen zur Verfügung stellt.

An vielen Standorten gibt es zudem separate Lehrbilsammlungen, für die eigenständige Infrastrukturen – abgekoppelt vom klinischen PACS bzw. den Forschungsarchiven – betrieben werden. Primäre Nutzer dieser Systeme sind Medizinstudenten und Dozenten.

## Rahmenbedingungen

In diesem Abschnitt werden verschiedene Rahmenbedingungen, die Einfluss auf die medizinische Bildgebung in der Forschung sowie auf das Design und den Einsatz von IT-Systemen in diesem Umfeld haben, kurz vorgestellt. Hierbei wird zwischen technischen, rechtlichen und finanziellen Rahmenbedingungen unterschieden.

### **Technische Rahmenbedingungen**

Die technischen Rahmenbedingungen werden durch Dateiformate und Kommunikationsstandards bestimmt. Während die Erhebung, Kommunikation, Auswertung und Archivierung der Originalbilder in der DICOM-Welt durch diesen weitverbreiteten, konzeptbedingt jedoch abschnittsweise heterogenen Standard vorgegeben werden, geht es außerhalb davon deutlich uneinheitlicher vor. In der Non-DICOM-Welt setzen die Hersteller der Geräte eine Reihe von proprietären Formaten und Kommunikationsprotokollen neben Standards, wie beispielsweise TIFF, ein. Einzelne Standorte sind deshalb dazu übergegangen, Konvertierungs- bzw. Wrapping-Techniken zu nutzen, um u. a. die digitale Pathologie DICOM-konform zu lösen.

Im weiteren Verlauf der Verarbeitungskette sind die technischen Rahmenbedingungen geprägt durch eine Vielzahl unterschiedlichster Softwaretools zur Verarbeitung, spezielle Datenformate und uneinheitliche Kommunikationsprotokolle bzw. Schnittstellen zu den Archivierungssystemen.

Neben der institutionsinternen Kommunikation stellt auch die Übermittlung der Daten in die Außenwelt, beispielsweise bei multizentrischen Forschungsprojekten, eine technische Rahmenbedingung dar. In vielen Fällen befinden sich die Archivsysteme, insbesondere das klinische PACS, in Netzsegmenten, aus denen ein direkter Versand in die Außenwelt aus Sicherheitsgründen nicht möglich ist. Stattdessen muss zunächst eine „Lücke“ zwischen dem Intranet und dem Internet überwunden werden, bevor die Datenübertragung durchgeführt werden kann.

## **Rechtliche Rahmenbedingungen**

Die rechtlichen Rahmenbedingungen sind durch das Bürgerliche Gesetzbuch (BGB – allgemeine Verjährungsfristen), die Röntgenverordnung (RöV – Aufbewahrungs- und Verjährungsfristen beim Einsatz von ionisierender Strahlung), die Datenschutzgesetze des Bundes und der Länder, die Landeskrankenhausgesetze – insbesondere was die Sekundärnutzung von Daten aus der Krankenversorgung angeht – sowie das Medizinproduktegesetz (Zertifizierung von Software und Geräten) und die Regeln zur guten wissenschaftlichen Praxis (Aufbewahrungsfristen für Daten, die Grundlage wissenschaftlicher Veröffentlichungen sind) vorgegeben. Im Kontext klinischer Studien müssen zusätzlich internationale Richtlinien, wie beispielsweise CFR 21 Part 11<sup>3</sup> der United States Food and Drug Administration (FDA), sowie die Good Clinical Practice (GCP<sup>4</sup>) der International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) berücksichtigt werden.

## **Finanzielle Rahmenbedingungen**

Die finanziellen Rahmenbedingungen gestalten sich sehr unterschiedlich: Die Kosten für die Bildgebung sind in klinischen Studien teilweise in den geförderten oder durch einen Sponsor finanzierten patientenbezogenen Fallpauschalen enthalten. Diese erweisen sich jedoch nicht in allen Fällen als kostendeckend. Hinzu kommen Abgrenzungsprobleme zur Bildgebung im Behandlungszusammenhang, bei der eine Kostenerstattung durch die Krankenkassen erfolgt. Die Kosten für die Infrastrukturen zur Verwaltung und Archivierung der Bilder werden dadurch jedoch nicht abgedeckt. An einzelnen Standorten gibt es Ansätze zur Grundfinanzierung von Infrastrukturen in Forschung und Lehre. Dennoch sollte in Förderanträgen darauf geachtet werden, Gelder für die Finanzierung der Infrastruktur mit zu beantragen, wenn dies seitens des Förderers ermöglicht wird. An dieser Stelle sei zusätzlich darauf hingewiesen, dass die notwendige Betriebsdauer der Infrastruktur, vor allem mit Blick auf gesetzliche Aufbewahrungs-

---

3 <http://www.fda.gov/RegulatoryInformation/Guidances/ucm125067.htm>

4 <http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html>

fristen und die für die wissenschaftliche Nachnutzbarkeit gewünschte Langzeitarchivierung, die eigentliche Projektlaufzeit in der Regel deutlich übersteigt. Daher ist eine Kofinanzierung aus Standortmitteln häufig unumgänglich.

## Aktueller Bestand und Erhebungsergebnisse

Die aktuelle Erhebung versucht zunächst einen Überblick über die verschiedenen bildgebenden Modalitäten, sowie die Daten, die sie produzieren und die damit verbundenen Metadatenmodelle zu geben. Anschließend werden einzelne Produkte als Vertreter der verschiedenen Kategorien von IT-Systemen in der forschungsbezogenen Bildgebung vorgestellt. Abschließend werden die Besonderheiten, die bei der Pseudonymisierung und dem Versand bzw. der Übertragung von Bilddaten zu beachten sind, aufgezeigt.

### Daten- und Metadatenformate

Wie bereits eingangs im Abschnitt zur Kategorisierung der IT-Landschaft erwähnt, kann bei den bildgebenden Modalitäten zwischen DICOM- und Non-DICOM-Geräten unterschieden werden. Erstere, beispielsweise CT, MRT, PET und Ultraschall, geben ihre Bilder im DICOM-Format aus. Auf der anderen Seite gibt es die Non-DICOM-Geräte, beispielsweise diverse Mikroskopie-Systeme, deren Hersteller überwiegend auf eigene, häufig TIFF-basierte, Formate setzen. Das Volumen eines Datensatzes variiert stark je nach Untersuchungsart und in Abhängigkeit von Geräteparametern, wie beispielsweise räumlicher und zeitlicher Auflösung sowie Farb- bzw. Grauwerttiefe. Die unterschiedlichen Welten haben auch einen direkten Einfluss auf Art und Format der Metadaten, die zusätzlich zu den eigentlichen Pixeldaten erfasst werden. Das DICOM-Format bietet neben allgemeinen Informationen, wie beispielsweise dem Zeitpunkt der Bildgebung, auch die Möglichkeit, eine Vielzahl von gerätespezifischen Parametern zu hinterlegen. Dafür wurden in spezialisierten Working Groups für die einzelnen Modalitäten sogenannte Service-Object Pair Klassen (SOP) entwickelt, die ausweisen, welche gerätespezifischen Parameter und Einstellungen unter welchen Tags und in welchem Format gespeichert werden sollen. Durch diese grundlegende Standardisierung ist gewährleistet, dass das Auslesen, Anzeigen und Interpretieren der Metadaten herstellerübergreifend weitestgehend einheitlich stattfinden kann. Allerdings hinkt die Entwicklung des Standards teilweise der realen technischen Entwicklung hinterher, sodass neue Messverfahren nicht immer korrekt in den Metadaten abgebildet werden können. Hinzu kommt, dass die Erfassung vieler Tags – insbesondere bei den ursprünglichen CT- und MRT-SOP-Klassen – optional ist und auch die Inhalte der Tags nicht immer dem Standard entsprechen. Mit der Verabschiedung neuer verbesserter MRT- und CT-

SOP-Klassen (Supplements 49<sup>5</sup> und 58<sup>6</sup>) in den Jahren 2002 und 2004 wurden einige dieser Punkte zwar adressiert und beispielsweise deutlich mehr Tags als verpflichtend eingestuft, allerdings sind die alten SOP-Klassen nach wie vor weit verbreitet. Darüber hinaus arbeiten einige weit verbreitete Verarbeitungswerkzeuge, beispielsweise FSL und SPM, mit den Formaten NIfTI oder Analyze. Diese Formate bieten gegenüber dem DICOM-Format Vorteile bei der Auswertung und Verarbeitung der Bilddaten: Durch den Verzicht auf Metadaten, die für die Kommunikation und Speicherung der Daten in einer DICOM-Umgebung notwendig wären, enthalten diese Formate deutlich weniger Overhead. Außerdem bietet das NIfTI-Format ein affines Koordinatensystem, welches die Verarbeitung erleichtert. Die notwendige Konvertierung der DICOM-Dateien führt jedoch zum Verlust der meisten Metadaten.

Dass der Verlust von Metadaten während der Verarbeitungskette, die medizinische Bilddaten im Forschungskontext durchlaufen, ein Problem für die Dokumentation der Provenienz und auch für das Teilen und Nachnutzen von Daten ist, hat unter anderem die International Neuroinformatics Coordination Facility erkannt und die Neuroimaging Data Sharing Task Force (NIDASH) gegründet. NIDASH entwickelt das grundsätzlich auf dem PROV-Modell des World Wide Web Consortiums (W3C) basierende Neuroimaging Data Model (NI-DM). Das Ziel von NI-DM ist es, den Austausch und die Veröffentlichung von Informationen über die Provenienz eines Datensatzes in der neurologischen Bildgebung zu ermöglichen und somit die Reproduzierbarkeit von wissenschaftlichen Ergebnissen zu fördern. Im Gegensatz zu beispielsweise DICOM und NIfTI soll NI-DM die gesamte Lebensspanne, von der Generierung über die Verarbeitung bis hin zur Auswertung und den damit verbundenen Formatwechseln abdecken [1]. Diese Provenance-Anforderungen werden nicht nur an Bilddaten gestellt, sondern an jegliche Daten, die möglicherweise an verschiedenen Standorten zu verschiedenen Zeitpunkten neu kombiniert bzw. weiter analysiert werden.<sup>7</sup>

In der Non-DICOM-Welt ist eine vergleichbare Art an herstellerübergreifender Standardisierung und daraus folgender Interoperabilität hingegen so gut wie gar nicht vorhanden. Zwar hinterlegen die Hersteller oftmals auch grundlegende technische Metadaten in ihren proprietären Datenformaten, allerdings können diese dann meistens nur von der hauseigenen Software interpretiert werden. Auch das Anzeigen und die Weiterverarbeitung der Bilder sind meistens nur eingeschränkt mit Produkten von Drittanbietern möglich. Im Bereich der Mikroskopie-Bildgebung gibt es mit Bio-Formats<sup>8</sup> eine Initiative,

---

5 [ftp://medical.nema.org/medical/dicom/final/sup49\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/sup49_ft.pdf)

6 [ftp://medical.nema.org/medical/dicom/final/sup58\\_ft2.pdf](ftp://medical.nema.org/medical/dicom/final/sup58_ft2.pdf)

7 Siehe hierzu auch das Kapitel zum Forschungsdatenmanagement in diesem Buch.

8 <http://www.openmicroscopy.org/site/support/bio-formats5.1/>

deren Ziel es ist, proprietäre Bildformate in das freie Open Microscopy Environment Model (OME) bzw. das Bildformat OME-TIFF umzuwandeln. Bio-Formats stellt hierfür eine Java-Bibliothek zur Verfügung, mit Hilfe derer viele proprietäre Bildformate gelesen und geschrieben werden können. Auch das Auslesen von Metadaten und die Konvertierung zu OME-XML werden für ausgewählte Formate unterstützt. Wie umfassend ein proprietäres Format unterstützt wird, hängt jedoch unter anderem von der Verfügbarkeit öffentlicher Dokumentation über dieses Format ab [2].

## Software und Systeme

Das Feld der Tools zur Bildverarbeitung ist von einer Vielzahl von Open-Source-Softwarestacks und einer sehr starken Spezialisierung auf einzelne Fachdisziplinen oder Methoden geprägt. Daher wird hier lediglich eine Auswahl präsentiert. Aus denselben Gründen wird auch von einer umfangreichen Analyse der einzelnen Produkte abgesehen. Die Gliederung der Produkte folgt der eingangs erstellten Kategorisierung gemäß Verwendungszweck:

1. Systeme zur Verwaltung, Archivierung und zum Austausch von Bildern im Forschungskontext
  - ▶ XNAT<sup>9</sup> (Washington University of St.Louis, MO, USA) ist eine webbasierte Open-Source Software zur Archivierung, Verwaltung und zum Austausch von medizinischen Bilddaten und assoziierten Daten im Forschungskontext. Sie wird sowohl für zentrale Forschungsdaten-Repositories, als auch für multizentrische Studien benutzt. XNAT kann u.a. mit Bilddaten im DICOM- oder Analyse-Format und mit allgemeinen Daten im XML-Format umgehen [3].
  - ▶ OMERO<sup>10</sup> (Open Microscopy Environment Consortium) ist eine Open-Source Client-Server Software zur Visualisierung, Verwaltung und Analyse von mikroskopischen Bilddaten. OMERO unterstützt viele proprietäre Formate über die Bio-Formats Library und bietet zudem noch ein eigenes offenes Bildformat (OME-TIFF) und Metadaten im OME-XML Format [4].
  - ▶ LORIS<sup>11</sup> (McGill University, Quebec, Canada) ist eine webbasierte Open-Source Software zur Verwaltung, Visualisierung und zum Austausch von medizinischen Bilddaten und assoziierten Verhaltens- und Genomdaten im Forschungskontext [5].

---

9 <http://www.xnat.org>

10 <http://www.openmicroscopy.org/site/products/omero>

11 <http://mcin-cnim.ca/neuroimagingtechnologies/loris/>

- ▶ MDPE<sup>12</sup> (Universitätsmedizin Mainz, Deutschland) – Medical Data and Picture Exchange. Eine webbasierte (Client-Server) Teleradiologieplattform mit der Möglichkeit, „Patientenakten“ in Projekten/Studien zu gruppieren. Unterstützt DICOM, Objekte aus der Strahlentherapie (RTSTRUCT, RTPLAN) und herkömmliche Bildformate (JPEG, PNG), sowie die Dokumentenformate Office und PDF.
- ▶ CTK<sup>13</sup> - The Common Toolkit. CTK ist eine Open-Source DICOM-Plattform, mit der es möglich ist, DICOM-Bilder aller Art zu speichern, anzuzeigen, zu senden und zu empfangen.
- ▶ DCMTK<sup>14</sup> – DICOM-Toolkit (Offis Oldenburg). DCMTK implementiert weite Teile des DICOM-Standards in Form von Anwendungen und Software-Bibliotheken und stellt diese als Open-Source zur Verfügung. In vielen Produkten werden DCMTK Bibliotheken zur Sicherstellung grundlegender DICOM-Funktionalitäten eingesetzt. Die bereitgestellten Anwendungen können eigenständig betrieben werden.
- ▶ TRIAL-Connect<sup>15</sup> (Telekom Healthcare Solutions). Eine webbasierte Forschungsplattform zur zentralen Verwaltung und Archivierung von Bilddaten in klinischen Studien und Forschungsprojekten. Zusätzlich ist ein umfangreicher Web-Viewer enthalten.
- ▶ Chili PACS<sup>16</sup> und Telemedizinakte (Chili GmbH Heidelberg/Dossenheim). Eine webbasierte Plattform zum institutionsübergreifenden Austausch sowie zur Verwaltung von radiologischen Bilddaten. Zusätzlich ist ein umfangreicher Web-Viewer enthalten.

## 2. Tools zur Verarbeitung und Auswertung (beispielhafte Auflistung)

- ▶ Freie bzw. Open-Source Tools: ITK<sup>17</sup>, MITK<sup>18</sup>, Fiji<sup>19</sup>, ImageJ<sup>20</sup>, FSL<sup>21</sup>, SPM<sup>22</sup>, FreeSurfer<sup>23</sup>, 3DSlicer<sup>24</sup>, MIPAV<sup>25</sup>
- ▶ Kommerzielle Tools: BrainVoyager<sup>26</sup>, CMR42<sup>27</sup>, Definiens<sup>28</sup>, MATLAB<sup>29</sup>, Imaris<sup>30</sup>

12 <http://www.unimedizin-mainz.de/imbei/informatik/projekte/aktuell/mdpe.html>

13 <http://www.commonstk.org>

14 <http://dicom.offis.de/dcmthk.php.de>

15 <http://www.telekom-healthcare.com/kliniken/studien-daten-verwalten-mit-trial-connect/1227258>

16 <http://www.chili-radiology.com>

17 <http://www.itk.org>

18 <http://www.mitk.org/>

19 <http://fiji.sc/Fiji>

20 <http://imagej.nih.gov/ij/>

21 <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>

22 <http://www.fil.ion.ucl.ac.uk/spm/>

23 <http://freesurfer.net/>

24 <http://www.slicer.org/>

25 <http://mipav.cit.nih.gov/>

26 <http://www.brainvoyager.com/>

27 <http://www.circlecvi.com/>

28 <http://www.tissuestudio.definiens.com>

29 <http://de.mathworks.com/products/matlab/>

30 <http://www.bitplane.com/imaris/imaris>

3. Produkte, die in der Lehre eingesetzt werden
  - ▶ Digitale Pathologielösungen (z. B. Olympus NIS<sup>31</sup> oder Leica Digital SlideBox<sup>32</sup>)
  - ▶ RSNA MIRC Teaching File System<sup>33</sup>
  - ▶ IHE TCE Selector<sup>34</sup>

## **Pseudonymisierung von Bilddaten**

Während der Pseudonymisierung<sup>35</sup> sollten Bilddaten weitestgehend von identifizierenden Informationen bereinigt werden. Insbesondere bei DICOM-Daten befinden sich häufig Informationen dieser Art in den Metadaten, sofern die Bilder in einem klinischen Kontext erzeugt worden sind. Die Inhalte der relevanten Tags müssen daher gelöscht oder pseudonymisiert werden, bevor die Daten weitergegeben werden dürfen. Die NEMA hat mit dem DICOM Supplement 142<sup>36</sup> einen Leitfaden für die Bereinigung von DICOM-Daten herausgegeben. Das Dokument spezifiziert sogenannte Application Level Confidentiality Profiles und führt ein Basisprofil mit verschiedenen Optionen ein. In der Praxis sollte der Einsatz eines solchen Profils unbedingt im Umfeld der späteren Anwendung getestet werden, da beispielsweise das Entfernen aller privaten Tags – was durchaus in einigen Profilen vorgesehen ist – in manchen Software-Umgebungen zu Problemen führen kann.

Auch beim Umgang mit Non-DICOM-Daten sollte unbedingt geprüft werden, ob identifizierende Informationen, beispielsweise die Patientenummer, in den Metadaten des Bildes gespeichert sind. Einen allgemeingültigen Leitfaden zur Bereinigung der Metadaten gibt es aufgrund der fehlenden Standardisierung der Dateiformate nicht. Häufig hilft jedoch ein Blick in die Dokumentation oder der enge Kontakt zum Hersteller der bildgebenden Modalität.

Zusätzlich zum Vorkommen von identifizierenden Merkmalen in den Metadaten einer Bilddatei können auch in den Pixeldaten kritische Informationen enthalten sein. Das können zum einen identifizierende Daten, wie beispielsweise der Name des Patienten, sein, die von der Modalität in das Bild eingebrannt worden sind und zum anderen können natürlich auch die Bildinformationen selber bereits ein hohes Re-Identifikationsrisiko bergen; beispielsweise bei CT- oder MRT-Aufnahmen vom Kopf einer Person.

---

31 <http://www.olympus-lifescience.com/en/microscopes/virtual/vs120/net-image-server-sql/>

32 <http://www.leicabiosystems.com/de/pathologische-bildgebung/aperio-epathology/integrieren/details/product/digital-slidebox/>

33 <https://www.rsna.org/tfs.aspx>

34 [http://mirwiki.rsna.org/index.php?title=TCE\\_Selector](http://mirwiki.rsna.org/index.php?title=TCE_Selector)

35 Allgemeine Hinweise zur Pseudonymisierung finden Sie im Kapitel Identitätsmanagement.

36 [ftp://medical.nema.org/medical/dicom/final/sup142\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/sup142_ft.pdf)

Das DICOM Supplement 142 adressiert grundsätzlich zwar auch diese beiden Fälle, überlässt die konkreten Methoden zur Entfernung dieser Informationen jedoch den Herstellern, die den Standard implementieren. Es existieren einige Techniken und Tools am Markt, die sich auf diese Anwendungsfälle spezialisiert haben und das Entfernen der Merkmale aus den Pixeldaten weitestgehend automatisieren [6].

### **Standortübergreifende Übertragung von Bilddaten**

Bei multizentrischen Forschungsprojekten werden Bilddaten häufig in zentralen Systemen gesammelt und verwaltet. Die Übertragung aus den erhebenden Einrichtungen in das zentrale Archiv stellt die Projekte jedoch mitunter vor einige Probleme. Insbesondere dann, wenn Modalitäten aus der Krankenversorgung genutzt werden, verhindert allein schon die vielerorts übliche Abschottung des Netzwerks eine direkte Übertragung der Daten. Bevor die Bilddaten also an das System des Forschungsprojektes übermittelt werden können, muss zunächst einmal die Lücke zwischen dem abgeschotteten lokalen Netzwerk und dem offenen Internet überwunden werden. Hierfür existieren Lösungen, wie beispielsweise der frei verfügbare Clinical Trials Processor (CTP)<sup>37</sup> der Radiological Society of North America (RSNA), aber auch kommerzielle Anbieter haben Systeme im Angebot, die – betrieben in einer demilitarisierten Zone – als Gateway für den Datenversand dienen können. Allerdings dürften die wenigsten Forschungsprojekte in der Lage sein, den finanziellen, personellen und organisatorischen Aufwand zu leisten, den die Installation und der Betrieb solcher Gateways an allen teilnehmenden Zentren bedeutet. Eine nennenswerte Ausnahme bildet hier die NAKO Gesundheitsstudie.<sup>38</sup> In vielen anderen Verbänden ist es jedoch gängige Praxis, die Trennung der Netze durch den Export der Daten auf einen mobilen Datenträger (CD, USB-Stick, externe Festplatte) zu überbrücken. Anschließend können die Daten dann entweder von einem Endpunkt im offenen Netzwerk hochgeladen oder direkt per Post auf dem Datenträger verschickt werden. Für letzteres gibt es sowohl Best-Practice Guidelines, als auch unterstützende Software, die im Rahmen des SWABIK-Projektes entwickelt worden sind.<sup>39</sup>

### **Aufbau von Lehrbilsammlungen in der Radiologie**

Lehrbilsammlungen sind ein wichtiger Bestandteil in der Ausbildung von Medizinstudierenden. Das Anlegen und Pflegen einer umfangreichen Lehrbilsammlung erfordert einigen Aufwand, um die richtigen Bilddaten zu selektieren, zu anonymisieren, mit klinischen Informationen anzureichern und zu Schlagwörtern zu versehen. Das

---

37 [http://mircewiki.rsna.org/index.php?title=CTP-The\\_RSNA\\_Clinical\\_Trial\\_Processor](http://mircewiki.rsna.org/index.php?title=CTP-The_RSNA_Clinical_Trial_Processor)

38 <http://nako.de/>

39 <http://swabik.offis.de/en/docs.html>

open-source Projekt MIRC (Medical Imaging Resource Community) der RSNA bietet mit dem Teaching File System (TFS<sup>40</sup>) ein Produkt mit sehr umfangreichen Möglichkeiten zum Anlegen und Nutzen einer Lehrbildersammlung. Da es die entsprechenden DICOM-Funktionalitäten unterstützt, lässt sich TFS sehr gut mit bestehenden PACS verbinden. Um die Auswahl und das Anreichern der Bilder mit den relevanten Informationen im Alltag zu erleichtern, wurde das IHE TCE Profil<sup>41</sup> entwickelt. Das TCE Profil definiert drei Aktoren: den Selector Actor, den Export Manager Actor und den Receiver Actor. Die letzten beiden werden bereits vom TFS implementiert. PACS, die IHE TCE unterstützen, setzen den Selector Actor um. Falls das eigene PACS IHE TCE nicht implementiert, bietet MIRC mit dem frei verfügbaren TCE Selector ein Werkzeug an, welches beispielsweise auf einer PACS Workstation installiert werden kann und eine grafische Oberfläche zur Auswahl der Bilder, zur Verschlagwortung und zum Versand an das TFS bereitstellt [7].

## Bewertung und Handlungsbedarf

Besonders auf dem Gebiet der Bilddatenverarbeitung unerfahrene Forschungsverbände neigen dazu, die Komplexität des Themas zu unterschätzen und auf Eigenentwicklungen bei der Infrastruktur zu setzen. Hiervon ist in Anbetracht der bereits verfügbaren, teilweise lizenzkostenfreien Lösungen jedoch abzuraten. Vorzugsweise sollte auf ein bestehendes System gesetzt und dieses an die eigenen Bedürfnisse angepasst werden. Die Wahl des Systems sollte dabei von den Anforderungen des Forschungsverbundes bezüglich der zu speichernden Daten, der Interoperabilität mit bestehenden Systemen und der Unterstützung von Verarbeitungsprozessen sowie des Qualitätsmanagements abhängig gemacht werden.

Handlungsbedarf sehen die Autoren insbesondere bei der einrichtungübergreifenden Kommunikation von Bilddaten, die in multizentrischen Projekten sowohl Forschungsverbände als auch teilnehmende Einrichtungen immer wieder vor Probleme stellt. Zurzeit mangelt es an standardisierten Lösungen zur Kommunikation zwischen klinischen Bildarchiven und Forschungssystemen unter Einbezug eines Pseudonymisierungsdienstes. Dies führt dazu, dass viele Forschungsverbände dieses Problem unter Einsatz eigener Ressourcen von neuem für sich lösen müssen.

In den Einrichtungen, die häufig Teil mehrerer Verbände sind, resultiert dies in vielen unterschiedlichen Prozessen für identische oder zumindest sehr ähnliche Aufgaben. Perspektivisch sollte die technologische Weiterentwicklung der Bildgebung in der Krankenversorgung und der Forschung daher besser als bisher verzahnt werden.

---

40 <https://www.rsna.org/tfs.aspx>

41 [http://wiki.ihe.net/index.php/Teaching\\_File\\_and\\_Clinical\\_Trial\\_Export](http://wiki.ihe.net/index.php/Teaching_File_and_Clinical_Trial_Export)

Eine weitere Standardisierung, über das bisherige Anwendungsspektrum von DICOM hinaus, erscheint im Kontext von Open Access und Datennachnutzung dringend notwendig. Insbesondere für die Dokumentation der Provenienz von Bild- und Ergebnisdaten werden Standards benötigt, um Nachvollziehbarkeit und Reproduzierbarkeit von Forschungsergebnissen zu verbessern. Hier stellt das in diesem Kapitel angesprochene Neuroimaging Data Model (NI-DM) einen vielversprechenden Ansatz dar. Es bleibt jedoch abzuwarten, ob sich NI-DM etablieren und ggf. auch in anderen Fachbereichen Anwendung finden wird.

Weiterer Handlungsbedarf besteht bei der Mikroskopie-Bildgebung. Trotz einiger Standardisierungsversuche, beispielsweise durch die DICOM Supplements 122<sup>42</sup> und 145<sup>43</sup>, arbeiten viele Systeme heute noch mit proprietären Formaten und Kommunikationsprotokollen. Dies führt zwangsläufig zu Problemen beim Mischbetrieb mit Geräten unterschiedlicher Hersteller und zu einem Lock-In-Effekt in den Einrichtungen. Das Open Microscopy Environment Consortium versucht zwar, diesen Problemen mit der Bio-Formats-Library, sowie dem OME-TIFF-Dateiformat entgegenzuwirken und bietet mit OMERO eine herstellerunabhängige Lösung an. Die nachträgliche Konvertierung zum OME-TIFF-Format birgt jedoch eigene Risiken, beispielsweise den Verlust von Metadaten. Wünschenswert wäre daher eine Standardisierung unter Einbeziehung der Gerätehersteller. Es bleibt abzuwarten, ob sich dieser Bereich weiterhin in eine eigene Richtung entwickelt, oder ob er perspektivisch doch noch Teil der DICOM-Welt wird.

## Literatur

1. Keator, D.B., Helmer, K., Steffener, J. et al., *Towards structured sharing of raw and derived neuroimaging data across existing resources*. Neuroimage, 2013. **82**: S. 647-661.
2. Linkert, M., Rueden, C.T., Allan, C. et al., *Metadata matters: access to image data in the real world*. J Cell Biol, 2010. **189**(5): S. 777-782.
3. Marcus, D.S., Olsen, T.R., Ramaratnam, M., Buckner, R.L., *The Extensible Neuroimaging Archive Toolkit: An Informatics Platform for Managing, Exploring, and Sharing Neuroimaging Data*. Neuroinformatics, 2007. **5**: S. 11 - 34.
4. Allan, C., Burel, J.M., Moore, J. et al., *OMERO: flexible, model-driven data management for experimental biology*. Nat Methods, 2012. **9**(3): S. 245-253.
5. Das, S., Zijdenbos, A.P., Harlap, J., Vins, D., Evans, A.C., *LORIS: a web-based data management system for multi-center studies*. Front Neuroinform, 2011. **5**: S. 37.

---

42 [ftp://medical.nema.org/medical/dicom/final/sup122\\_ft2.pdf](ftp://medical.nema.org/medical/dicom/final/sup122_ft2.pdf)

43 [ftp://medical.nema.org/medical/dicom/final/sup145\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/sup145_ft.pdf)

6. Bischoff-Grethe, A., Ozyurt, I.B., Busa, E. et al., *A technique for the deidentification of structural brain MR images*. Hum Brain Mapp, 2007. **28**(9): S. 892-903.
7. Mildenerger, P., Brüggemann, K., Rösner, F., Koch, K., Ahlers, C., *PACS infrastructure supporting e-learning*. Eur J Radiol, 2011. **78**(2): S. 234-238.



# Biobanken

*Sara Y. Nußbeck, Hans-Ulrich Prokosch*

## Anwendungsbereich

Biomaterialbanken, kurz Biobanken spielen im Bereich der biomedizinischen Forschung eine immer größere Rolle [1]. Die Verwaltung einer großen Anzahl einzelner Proben, wie Blut oder Gewebe, und die Verknüpfung mit den entsprechenden phänotypischen Informationen des Spenders sind ohne eine professionelle IT-Unterstützung nicht mehr leistbar [2, 3]. In den Aufbau großer zentralisierter Biobanken hat das Bundesministerium für Bildung und Forschung (BMBF) von 2011 bis 2016 über 20 Millionen Euro investiert.<sup>1</sup> Ende 2015 erfolgte die Ausschreibung zur „Ertüchtigung deutscher Biobanken für die Anbindung an BBMRI“, die sogenannte German Biobank Alliance (GBA). In diesem Rahmen sollen von 2017-2019 zehn Biobankstandorte unter der Koordination des German Biobank Node (GBN) u. a. IT Standards und Werkzeuge für die deutsche Biobanking Community entwickeln und testen. Parallel dazu veröffentlichte das BMBF die Medizininformatikinitiative<sup>2</sup>, in deren Rahmen Konsortien gebildet werden sollen, um gemeinsame Datenintegrationszentren zu konzipieren und später aufzubauen. Da in diesem Kontext jeweils Anwendungsfälle beschrieben werden sollen, anhand derer letztlich auch der Nutzen der Datenintegrationszentren gezeigt werden kann, werden Biobanken mit ihrer notwendigerweise engen Verknüpfung mit den phänotypischen Daten hier ebenfalls eine Rolle spielen.

In diesem Kapitel über IT-Systeme zur Unterstützung des Biobanking (Biobankmanagementsysteme) sind primär Softwareprodukte berücksichtigt, die zur Verwaltung der Biomaterialien selbst, sowie der entsprechenden Lagerressourcen dienen. Dazu gehört vor allem die Dokumentation der Lager- und Qualitätsparameter für jede einzelne Probe. Die Dokumentation von phänotypischen Probandendaten, sowie der aus dem Biomaterial zu einer bestimmten Forschungsfragestellung generierten Analysedaten, erfolgt typischerweise in anderen Softwarelösungen, die hier nicht betrachtet werden. Unter anderem ist dies auch auf Empfehlungen zur informationellen Gewaltenteilung aus den TMF-Datenschutzkonzepten [4] zurückzuführen, die allerdings nicht zwangsweise die Nutzung unterschiedlicher Softwaresysteme vorschreiben. Da für die Recherchierbarkeit von Biomaterial in Forschungskonsortien aber in der Regel Standort-bezogen eine Integration verschiedener Datenarten in Forschungs-Data-Warehouse-Systemen erfolgt, und diese wiederum deutschlandweit (vgl. hierzu die aktuellen Projektarbeiten

---

1 Siehe <http://www.gesundheitsforschung-bmbf.de/de/2638.php>.

2 Siehe <http://www.gesundheitsforschung-bmbf.de/de/medizininformatik.php>.

von GBN und GBA), aber auch europaweit miteinander zu vernetzen sind (vgl. BBMRI-ERIC Projekt ADOPT sowie die Entwicklung der Common Services IT), wird auf entsprechende aktuelle Entwicklungen in diesem Kapitel zumindest grob eingegangen.

Prinzipiell sollten in einem Biobankmanagementsystem alle Daten erfasst werden, die mit dem Lebenszyklus einer Probe zusammenhängen. Das beginnt mit der Entnahme der Probe, ihrem Transport zur Biobank, der Probenverarbeitung, dem Einfrieren und Lagern und reicht bis hin zum Auftauen, der Verarbeitung der Probe bezüglich des angefragten Materials, der Qualitätsbestimmung und dem Versand. Letztendlich sollte im Sinne der vollständigen Lebenszyklus-Dokumentation einer Probe auch nachvollziehbar sein, für welche Art von Analyse diese (bzw. Aliquots aus der Probe) an welche Institution abgegeben wurde, welche Art von Analyseergebnissen gewonnen wurden und wo diese Ergebnisse abgelegt sind.

In den letzten Jahren wurde im Fachgebiet „Biospecimen Science“ erforscht, welche Parameter Aussagen über die Qualität einer Biomaterialprobe zulassen. So wurde beispielsweise auch der Effekt präanalytischer Variablen auf die Probenqualität untersucht. Es gilt heute deswegen als „good scientific practice“, dass der komplette Lebenszyklus einer Probe genauestens dokumentiert wird [5-7]. Darüber hinaus muss eine Biomaterialverwaltungssoftware großer Biobanken die Anbindung und Ansteuerung von Lagerrobotern, Aliquotierrobotern und verschiedenen Laborgeräten gewährleisten, um ein effizientes Proben-Management, mit automatischer Dokumentation aller Verarbeitungsschritte, zu ermöglichen.

Obwohl im Bereich der Biobanken meist zwischen „liquid biobanking“ und „solid biobanking“ unterschieden wird, ist diese Unterscheidung im Bereich der Softwareprodukte eher irrelevant, auch wenn sich einige Workflows und Teile der zu dokumentierenden Daten unterscheiden. Die Bandbreite der Softwarelösungen in diesem Bereich ist jedoch sehr groß und reicht von Excel-sheets und eigenentwickelten Datenbanksystemen über Open-Source-Produkte bis hin zu kommerziellen Produkten. Mittlerweile ist aber ein deutlicher Trend hin zur Etablierung kommerzieller Biobankmanagementsysteme zu erkennen. Der 2009 im Rahmen des IT-Strategie-Projekts der TMF entwickelte Anforderungskatalog [3] hat auch heute noch seine Gültigkeit. Von den in der 2009 durchgeführten Marktanalyse betrachteten 21 Softwarelösungen sind in Deutschland nach unserem Kenntnisstand heute nur noch Starlims (Abbot Informatics) und in wenigen Installationen Nautilus (ThermoFisher Scientific) vertreten. Für das 3. Nationale Biobankensymposium 2014 in Berlin wurde das aktuelle Angebot an Softwarelösungen in Deutschland abgefragt und eine Übersicht bereitgestellt. Zusätzlich wurden Anforderungen beschrieben, welche bei der Auswahl einer Biomaterialverwaltung

betrachtet werden sollten [8]. Fast alle Biobanken haben erkannt, dass Dokumentation und Verwaltung von Biomaterialien heute eine professionelle IT-Infrastruktur erfordern und dass dafür ein standortübergreifender Erfahrungsaustausch unabdingbar ist. Die TMF hat mit ihren Arbeitsgruppen „Biomaterialbanken“, „Datenschutz“ sowie „IT-Infrastrukturen und Qualitätsmanagement“ eine Plattform für den Informations- und Erfahrungsaustausch innerhalb ihrer Mitgliedschaft und zunehmend auch darüber hinaus etabliert und auch durch ihre Publikationen [4, 9-11] dazu beigetragen, dass Forschungsverbände auf diesem Wissen aufbauen können. Darüber hinaus wurde 2013 mit dem GBN ein zentraler Anlaufpunkt zum Thema Biobanking für Deutschland etabliert<sup>3</sup>, der wiederum als Austauschplattform auf Arbeitsebene die bewährte Struktur der AG Biomaterialbanken der TMF einbindet. Neben dem Qualitätsmanagement und der Öffentlichkeitsarbeit stellen IT-Konzepte die zentralen Arbeitspakete des GBN dar. Gleichzeitig übernimmt die Geschäftsstelle des GBN die Rolle des deutschen Partners (BBMRI.DE) im Kontext der europäischen Biobankenvernetzung (BBMRI-ERIC) [12] und die Koordination der Vernetzung deutscher Biobanken im Rahmen der GBA.

## Betreiber, Anwender, Nutzer

Eine „Biobank“ ist definiert als eine Sammlung von Biomaterial von Menschen, Tieren, Pflanzen oder Mikroben und assoziierten Daten, wobei solche Sammlungen nach professionellen Standards verwaltet werden müssen [13]. Je nach organisatorischem Aufbau und Größe der Biobank können die Betreiber, Anwender und Nutzer eines Biobankmanagementsystems stark variieren.

Bei größeren Biobanken an zentralen Standorten wie zum Beispiel Universitätskliniken wird die Software meist vom lokalen Rechenzentrum oder einem Institut für Medizinische Informatik betrieben. Bei kleineren Biobanken ist der Betreiber häufig die betreibende Institution mit einem eigenen IT-Verantwortlichen (so werden Softwareanwendungen für Gewebebanken häufig von IT-Administratoren innerhalb der Pathologie betreut).

Gibt es mehrere Biobanken in einem Forschungsverbund, so stellt entweder einer der beteiligten Standorte die zentrale IT-Infrastruktur für das Biobanking im gesamten Forschungsverbund, oder aber die lokal bereits existierenden professionellen Biobankmanagementsysteme werden jeweils dezentral weiter genutzt [2, 14]. Insgesamt stellt sich in einem Verbund immer die Frage, wie eine alle „Teilbiobanken“ übergreifende Abfrage-/Retrievalmöglichkeit von Proben realisiert werden kann. Für diesen Zweck

---

<sup>3</sup> <http://bbmri.de>

sollte ein Werkzeug zur standortübergreifenden Zusammenführung oder zur verteilten Abfrage von Proben- und Phänotypdaten geschaffen werden. Die jeweils an den einzelnen Standorten vorhandenen, etablierten Softwarelösungen zum Biobankmanagement können unabhängig davon zum jeweiligen Proben-/Lagerortmanagement weiter verwendet werden.

Die erste Stufe zur „Vernetzung“ mehrerer Biobanken ist in der Praxis immer der sogenannte Katalog (Catalogue bzw. zukünftig im europäischen Kontext als Directory bezeichnet). Beispielhaft kann hierzu auf die von Kuhn und Kollegen beschriebenen Arbeiten zur Vernetzung von Biobanken großer europäischer Kohorten (BBMRI-LPC) verwiesen werden [15].

Die Charakterisierung der verschiedenen Nutzer der Software ist davon abhängig, wie die Biobank organisatorisch aufgebaut und etabliert ist. Generell können zwei Nutzertypen unterschieden werden: Nutzer, welche die probenbeschreibenden Daten eingeben, und Nutzer, die Abfragen ausführen, bzw. Proben aus der Biobank herausnehmen wollen. Zu den Dateneingebnern gehören die am Prozess der Probengewinnung beteiligten Personen, z. B. OP-Personal, Pflegepersonal und Studienassistenten (Probenentnahme), medizinische-technische Assistenten (Probenverarbeitung und Einlagerung) und Ärzte bzw. wissenschaftliches Personal (Qualitätssicherung). Ein höherer Automatisierungsgrad der Biobank bei der Dateneingabe führt dabei zu weniger manuellen Eingaben und damit zu einer geringeren Fehlerrate. Hierfür müssen jedoch die Schnittstellen zu den entsprechenden Geräten vorhanden sein. Nutzer, die Biomaterial anfordern, dürfen, je nach den Policies der Biobank, selber in der Software nach Proben suchen oder geben eine Anfrage an die Biobank in Auftrag. Danach wird manuell durch eine verantwortliche Person oder automatisiert nach den Proben mit den entsprechenden Eigenschaften gesucht. Ist die Biobank kleiner, so besitzen in der Regel wenige Personen jeweils mehrere Rollen, d. h. die dateneingebende und probenherausgebende Person sind beispielsweise dieselbe.

## Rahmenbedingungen

**Rechtlich:** Die rechtlichen Rahmenbedingungen, unter denen Biobanken agieren und Softwareanwendungen betrieben werden dürfen, sind in der Regel, je nach Trägerschaft und Bundesland, durch die jeweiligen Landesdatenschutzgesetze oder das Bundesdatenschutzgesetz vorgegeben. Für Biobanken, die im Kontext der Krankenversorgung etabliert werden, gilt darüber hinaus in den meisten Bundesländern das jeweilige Landeskrankenhausgesetz. Für eine ausführliche Betrachtung der rechtlichen Rahmenbedingungen sei auf Simon et al. [11] verwiesen. Grundsätze für eine datenschutzgerechte

IT-Unterstützung im Biobanking sind in der Regel die Pseudonymisierung personenbezogener Daten, die getrennte Haltung von Probandendaten und Probanden sowie eine entsprechende Aufteilung der Datenhoheit. Da Biomaterial meistens genetische Informationen enthält, die unter Umständen einen Rückschluss auf die Identität und ggf. die Krankheit des Spenders zulassen, sind Probandendaten und auch Analyseergebnisse, die aus den Proben im Nachhinein entstehen, als schützenswert einzustufen. Die Optionen zur Detailumsetzung einer IT-Lösung hängen von der jeweils gewählten Modellvariante einer Biomaterialbank ab und sind im Datenschutzkonzept der TMF für Biomaterialbanken [16] und im Datenschutzleitfaden der TMF [4] beschrieben.

Für die Verwendung von Biomaterial über die primäre Diagnostik im Rahmen der Krankenversorgung hinaus oder die zusätzliche Entnahme von Biomaterial im Rahmen eines Forschungsprojekts muss jeweils die unterschriebene Einwilligungserklärung des Spenders vorliegen. Diese muss nicht zwangsweise in einer Biomaterialverwaltungssoftware dokumentiert sein. Allerdings ist, gerade im Hinblick auf die aktuellen GBN-/BBMRI-Aktivitäten zur standortübergreifenden Recherche verfügbarer Biomaterialien, die Etablierung eines IT-basierten Einwilligungsmanagements zu empfehlen.

Es gibt darüber hinaus nationale und internationale Richtlinien und Empfehlungen in Bezug auf den Betrieb von Biobanken. Dazu gehören die folgenden Dokumente:

- ▶ Die Stellungnahme des Deutschen Ethikrats zu Humanbiobanken für die Forschung [17]
- ▶ Best Practices for Repositories: Collection, Storage, Retrieval and Distribution of Biological Materials for Research, herausgegeben von der International Society of Environmental and Biological Repositories (ISBER) [18]
- ▶ NCI Best Practices for Biospecimen Resources from the Office of Biorepositories and Biospecimen Research, herausgegeben vom National Cancer Institute (NCI) der US National Institutes of Health [19]
- ▶ OECD Best Practice Guidelines for Biological Resource Centres, veröffentlicht von der Organisation for Economic Co-operation and Development (OECD) [20]
- ▶ OECD Guidelines on Human Biobanks and Genetic Research Databases, ebenfalls veröffentlicht von der OECD [21]
- ▶ Die Empfehlung Rec (2006) 4 on research on biological materials of human origin des Ministerkomitees des Europarates [22]
- ▶ Die Deklaration des Weltärztebundes zu Ethischen Grundsätzen für die medizinische Forschung am Menschen, „Declaration of Helsinki“ [23]
- ▶ Die Deklaration von Taipei des Weltärztebundes zu ethischen Aspekten bei Gesundheitsdatenbanken und Biobanken [24]

Im Zuge der wachsenden Bedeutung von Biobanken für die Forschung wird seit einigen Jahren auch intensiv über die Notwendigkeit eines Biobankengesetzes [z. B. 25] oder eines gesetzlich verankerten Biobankgeheimnisses, wie vom Deutschen Ethikrat vorgeschlagen [17], diskutiert.<sup>4</sup> Unabhängig von der politischen und gesetzgeberischen Umsetzbarkeit eines solchen Vorhabens ist aus der Sicht der Forschung zu fordern, dass die Vorteile, wie z. B. die gewonnene Rechtssicherheit, sorgfältig gegen ebenfalls mögliche Nachteile durch die Verringerung der Flexibilität der Rahmenbedingungen – davon könnten gerade kleinere oder auch innovative Biobankinfrastrukturen betroffen sein – abgewogen werden müssen [vergl. 26].

**Organisatorisch:** Organisatorisch können Biobanken gemäß einer der drei von Pomerening [10] beschriebenen Modellvarianten betrieben werden. Dazu gehören

- ▶ Biobanken als Teil einer Klinik oder medizinischen Einrichtung,
- ▶ eigenständige Biobanken und
- ▶ Biobanken in einem Forschungsnetz.

Für den Einsatz entsprechender IT-Lösungen ist dabei relevant, ob Biomaterialien und Phänotypdaten im Versorgungskontext (z. B. in den Biobanken der Comprehensive Cancer Center) oder im reinen Studienkontext (z. B. in den Kompetenznetzen in der Medizin bzw. den Deutschen Zentren der Gesundheitsforschung) erhoben werden. Für eine effiziente Softwarenutzung ist es auf jeden Fall zu empfehlen, an einem Universitätsklinikumsstandort nicht für jede kleine Biobank voneinander unabhängige Biobankmanagementsysteme zu betreiben, sondern vielmehr ein zentrales, mandantenfähiges System zu etablieren, in dem sowohl eine zentrale Biobankstruktur als auch dezentral verteilte Biobanken abgebildet werden können. Da entsprechende Systeme an vielen deutschen Universitätsklinika schon etabliert sind bzw. gerade etabliert werden, sollten auch für studienbezogene Biobanken (egal ob mono- oder multizentrische) keine eigenen speziellen Biobankmanagementsysteme aufgebaut, sondern die entsprechenden Anforderungen im Rahmen eines bereits etablierten mandantenfähigen Systems unterstützt werden.

**Standardisierung:** Für den Fall, dass Biomaterialien zwischen verschiedenen Biobanken ausgetauscht bzw. für übergreifende Forschungsprojekte gemeinsam genutzt werden sollen, ist es ratsam, sowohl bei der Dokumentation der probenbeschreibenden Qualitätssicherungsdaten als auch der Phänotypdaten sowie bei den Metadaten zur Charakterisierung von Biobanken auf entsprechende Standards zu setzen. Beispielhaft

---

<sup>4</sup> siehe z. B. das Expertengespräch des Deutschen Ethikrats und der TMF vom 8.4.2011 (<http://www.tmf-ev.de/News/886>)

seien hier SPREC [5, 6], BRISQ [27], MIABIS [28, 29] sowie der ADT-Datensatz in den Comprehensive Cancer Center genannt. Im nicht humanen Bereich sind die entsprechenden Standards der „environmental“ SPREC [30] und der GGBN Standard [31]. Um flexibel auf unterschiedliche Kodierungen und Standards an verschiedenen Standorten reagieren zu können, sollten insbesondere standardisierte und automatisiert auswertbare Metadatenstandards genutzt werden.<sup>5</sup> Aktuell widmet sich die MIABIS Arbeitsgruppe von BBMRI generisch dem Thema der Datenharmonisierung zwischen Biobanken.

Zudem gibt es mit dem ISO TC 276 WG 2 Anstrengungen, einen internationalen Standard für Biobanken einzuführen. U. a. geht es beim Unterthema IT um das Datenmanagement von klinischen Daten und Biomaterialdaten sowie Interoperabilität und Authentifikationsverfahren. Mit der Fertigstellung der Norm wird für 2019 gerechnet [32].

## Kategorisierung der IT-Landschaft

Generell können Softwarelösungen für Biobanken unterschieden werden in Eigenentwicklungen, Adaptierungen von Fremdsoftware, Open-Source-Produkte und kommerzielle Produkte. Die Vor- und Nachteile der einzelnen Lösungen sollen hier nicht weiter erläutert werden, da diese für alle Einsatzbereiche, nicht nur Biobanken, nahezu identisch sind.<sup>6</sup> Die meisten Produkte zum Management von Biobanken stammen allerdings aus dem Bereich der Labor-Informationen-Management-Systeme (LIMS) oder aus dem Anwendungsfeld der Pathologie. Der wichtigste Aspekt einer Biobank ist die Verknüpfung der Biomaterialproben mit ihren Annotationsdaten. Dies sind die phänotypischen und klinischen Verlaufsdaten (inkl. Bilddaten) der Probenspender sowie evtl. auch Ergebnisse aus molekularen oder genetischen Analysen. Wenn Biomaterialien und zugehörige Daten in der Krankenversorgung erhoben werden, sind Schnittstellen zwischen dem Biobankmanagementsystem und den entsprechenden Abteilungssystemen des Krankenhauses notwendig, um manuelle, zeitintensive und fehleranfällige Datenübertragungen zu vermeiden. Hierzu gehören je nach vorhandener Infrastruktur das Klinische Arbeitsplatzsystem (KAS) bzw. die elektronische Patientenakte (ePA), das Laborinformationssystem (LIS), das Pathologiesystem und das OP-System. Da Gewebebanken (und Hirnbanken) in der Regel nur im Kontext der Krankenversorgung etabliert und meist von Pathologieinstituten bzw. Neuropathologien betrieben werden, ist für diese eine Schnittstelle zwischen dem (Neuro-) Pathologieinformationssystem und dem Biobankmanagementsystem von grundlegender Bedeutung. Wenn Blut- oder Urinproben aus dem Versorgungskontext in einer Biobank verwaltet werden, sollten auch Daten aus den Laborinformationssystemen übernommen werden können. Die

---

<sup>5</sup> Siehe Kapitel zu Metadaten und Metadatenrepositories im vorliegenden Report.

<sup>6</sup> Vgl. hierzu den entsprechenden Abschnitt im Kapitel zu klinischen Studien im vorliegenden Report.

IT-Infrastruktur einer Biobank umfasst zur Verknüpfung der Proben mit strukturierten Phänotypdaten in der Regel auch das Clinical Data Warehouse (CDW). In einem Comprehensive Cancer Center ist darüber hinaus oft auch das Dokumentationssystem des Klinischen Krebsregisters in diese IT-Architektur zu integrieren. Zusätzlich ist es unter Umständen sinnvoll, das Patientendatenverwaltungssystem des Krankenhauses so einzubinden, dass von dort zur Generierung einheitlicher Pseudonyme automatisiert die typischerweise von der elektronischen Gesundheitskarte eingelesenen Patientenstammdaten genutzt werden können.

Handelt es sich bei der Biobank um eine studienbasierte Biobank in der Forschung, so werden Schnittstellen zum ID-Management, zur Studiendatenbank und einer Forschungsdatenbank für die Abfrage, Probenrecherche und Kohortenidentifikation benötigt.

Beispielhafte IT-Architekturen, die an Universitätskliniken zur Unterstützung verschiedener Forschungsszenarien eingesetzt werden können, wurden von Demiroglu et al. [2] und Prokosch et al. [33] beschrieben.

## Aktueller Stand

Die im Weiteren aufgeführten Angaben zum aktuellen Stand des Einsatzes von Biomaterialverwaltungssoftware beruhen auf Informationsrecherchen, die zum einen (2014/2015) über den GBN bei den deutschen BMBF-geförderten zentralisierten Biobanken (cBMBs), den Deutschen Zentren der Gesundheitsforschung und der NAKO Gesundheitsstudie durchgeführt wurden, sowie einer im November 2015 unter den IT-Leitern der deutschen Universitätsklinika durchgeführten und im Oktober 2016 aktualisierten Umfrage. Für eine Übersicht zum internationalen Stand, die auf einer Online-Umfrage der ISBER Working Group (WG) Informatics von 2012 beruht, an der sich 71 Personen beteiligten, sei auf [34] verwiesen.

In den frühen Jahren des Biobanking wurden in Deutschland kaum kommerzielle Biobankmanagementsysteme eingesetzt. Auf Basis eines der Abschlussberichte zum IT-Strategie-Projekt der TMF zur Analyse der derzeit in Deutschland verfügbaren IT-Werkzeuge zur Unterstützung des Managements von Biomaterialbanken [35] begannen in 2010 die ersten deutschen Biobanken mit der Einführung kommerzieller Biobankmanagementsysteme. In 2014 hatte sich die Situation bereits deutlich verändert. Bei einer im September/Oktober 2014 durchgeführten Umfrage im Rahmen der AG Biobanken und im Deutschen Biobankenregister gab es 30 Rückmeldungen. Diese lieferten Informationen sowohl zu zentralisierten Biobanken für ganze Universitätsstandorte als auch zu Biobanken einzelner klinischer Institute oder projektbasierten

Biobanken. Insgesamt gaben im Herbst 2014 etwas über 60% der befragten Biobanken an, für die Verwaltung der Biomaterialien auf ein kommerzielles, professionelles Biobankmanagementsystem zu setzen. Von den 36%, die zu diesem Zeitpunkt nicht auf ein solches kommerzielles Biobankmanagementsystem, sondern auf Eigenlösungen oder adaptierte Softwarelösungen setzten, war bei einem Drittel ein Wechsel zu einem kommerziellen Produkt geplant.

## **Universitätsklinika**

Ende 2015 wurde eine neuerliche Umfrage (dieses Mal unter allen deutschen Universitätsklinika; n = 32 Antworten) durchgeführt. Standorte, die 2015 angaben, in 2016 ein kommerzielles Produkt erwerben und einführen zu wollen, wurden im Oktober 2016 nochmals befragt. Insgesamt zeigen diese Umfragen, dass sich die meisten universitären Standorte mittlerweile für die Nutzung oder baldige Einführung eines kommerziellen Biobankmanagementsystems entschieden haben (22 bereits in der Produktivphase; 6 weitere Standorte planen die Einführung eines kommerziellen Produkts Ende 2016/2017).

Zwei Produkte haben sich in Deutschland als Marktführer etabliert. Es sind dies die Systeme Starlims (Abbott Informatics Germany GmbH) mit derzeit 6 Installationen (erste Starlims-Einführungen stammen bereits aus den Jahren 2010/2011), sowie das Produkt CentraXX (KAIROS) mit derzeit 13 Installationen. Die meisten CentraXX-Einführungen wurden im Jahr 2014 initiiert, weitere Einführungsprojekte der CentraXX-Lösung gab es in 2015 und 2016. Weitere eingesetzte kommerzielle Produkte (für die derzeit noch keine Ablösung geplant ist) sind Freezerworks, GenoMatch/SAMVentry und Nautilus mit jeweils einer Installation.

An 5 Universitätskliniken werden derzeit noch verschiedene Eigenentwicklungen genutzt, wobei die Ablösung durch ein kommerzielles Produkt für die Zukunft bei den meisten Standorten angedacht ist. An 2 Standorten ist bisher weder ein Biobankmanagementsystem im Einsatz, noch ist derzeit eine Produktbeschaffung geplant.

Da mit dieser Umfrage alle deutschen Universitätsklinika angesprochen waren, umfassen diese Ergebnisse auch die Standorte zentralisierter Biobanken (cBMBs). Eine gesonderte Betrachtung erübrigt sich somit.

## **Deutsche Zentren der Gesundheitsforschung**

Anders ist es bei den Deutschen Zentren der Gesundheitsforschung, die jeweils aus einer Vielzahl von Standorten und Partnerinstitutionen bestehen. An einem solchen

Zentrum sind in der Regel mehrere Universitätskliniken beteiligt. Um Synergieeffekte zu heben, sollte ein solches, mehrere Standorte umfassendes Gesundheitsforschungszentrum somit kein neues zusätzliches Biobankmanagementsystem etablieren, sondern auf dem an einem beteiligten Universitätsklinikum etablierten System aufsetzen. Die Verwaltung der Proben und Daten des Gesundheitsforschungszentrums kann dann als dedizierter eigener Mandant innerhalb einer solchen Installation etabliert werden.

Diese Vorgehensweise wurde z. B. im Deutschen Konsortium für Translationale Krebsforschung (DKTK) gewählt. Für das reine Biobankmanagement verwendet jeder der Standorte das bisher schon lokal etablierte IT-System weiter. Zum Zweck der Standort-übergreifenden Vernetzung kommt das in [36] beschriebene Brückenkopf-Konzept (mit einer „Light-Variante“ von CentraXX als „Brückenkopf“) zum Einsatz. Basis des Brückenkopfmodells ist die Annahme, dass Forscher aus Gründen des wissenschaftlichen Wettbewerbs Vorbehalte haben, Proben und Daten jedermann offenzulegen bzw. herauszugeben. Deswegen wurde eine Lösung etabliert, bei der jeder Forscher die Hoheit über seine Daten und Proben behält. Keine Daten und auch keine aggregierten Zahlen verlassen das Haus, solange das nicht ein lokal Verantwortlicher freigegeben hat. Bis zum Sommer 2015 konnten auf diese Art die klinischen Daten (entspricht größtenteils dem ADT-Tumordatensatz) von 9 der 11 Standorte bereits mittels „dezentraler Suche“ abfragbar gemacht werden.

Das Deutsche Zentrum für Herz-Kreislauf-Forschung (DZHK) setzt hingegen ein Basis-Biobanking um, d. h. von jedem Teilnehmer in einer DZHK-Studie wird eine bestimmte Menge an Biomaterial gesammelt. Die Prozesse zur Verwaltung der Biomaterialien sollen zukünftig durch ein zentrales Laborinformationssystem (DZHK-LIMS; die Entscheidung ist hier für CentraXX gefallen) IT-seitig unterstützt werden. Anfang Oktober 2016 ist der erste Standort produktiv gegangen. Solange das zentrale DZHK-LIMS noch nicht überall implementiert ist, wird eine abgestimmte Übergangslösung für das Biobanking in den ersten Studien umgesetzt. DZHK-Studienzentren, die noch kein Biobankmanagementsystem besitzen, sollen ihr Stellplatzmanagement der Biomaterialien über das noch zu beschaffende DZHK-LIMS abwickeln können. DZHK-Zentren, die bereits ein kommerzielles oder eigenentwickeltes LIMS in Einsatz haben, sollen über Schnittstellen an das zentrale DZHK-LIMS angebunden werden.

Im Deutschen Zentrum für Infektionsforschung (DZIF) erfolgt die Lagerung von Bioproben dezentral an allen Standorten und deren Verwaltung lokal mit den jeweils etablierten Biobankmanagementsystemen. Für retrospektive Abfragen ist ein DZIF-Katalog mit aggregierten Daten aus den dezentralen Biobanken und Sammlungen vorhanden. Für vom DZIF initiierte bzw. koordinierte Probensammlungen wird derzeit ein zentrales

Bioprobenregister (ZBR) implementiert, in dem ein gemeinsamer Minimaldatensatz auf Einzelprobenebene dokumentiert wird. Dieser liefert Informationen hinsichtlich Probenanzahl, Probenart, Lagerung, Präanalytik und relevante klinischen Daten, die eine gezielte DZIF-weite Probensuche durch Wissenschaftler ermöglichen. Außerdem ist so ein stets aktueller Überblick über Art, Menge und Verfügbarkeit der Bioproben an allen Standorten gegeben. Die Daten werden dabei weiter dezentral dokumentiert und über einen Extraktions-, Transformations- und Lade-Prozess (ETL) in das ZBR übertragen, das – gemäß der Definition von Lablans et al. [37] – einen zentralen Suchdienst zur Vernetzung der DZIF Biobanken realisiert. Die Implementierung des ZBR erfolgt zusammen mit der Firma Kairos auf Basis des Produkts CentraXX.

Das zugehörige Datenschutzkonzept sieht unter anderem vor, dass die Daten aus den einzelnen Quellsystemen im ETL-Prozess entsprechend den Vorgaben des Minimaldatensatzes aufbereitet und im Informationsvolumen reduziert werden. Die Registrierung der Probanden erfolgt dezentral in den teilnehmenden Institutionen und der Datenexport aus den Biobanken findet ausschließlich unter Verwendung von Pseudonymen statt. Somit werden keine unmittelbar personenidentifizierenden Informationen ans ZBR übermittelt. Die Datenübermittlung erfolgt nur nach erfolgreicher Authentifizierung des Absenders und innerhalb eines verschlüsselten Übertragungskanal. Der Datenaustausch wird uni-direktional durchgeführt, das bedeutet, dass keine Rückübermittlung der im ZBR gesammelten Daten an die Standorte erfolgt. Die Bearbeitung von Suchanfragen erfolgt zunächst ausschließlich über die das ZBR verwaltende Einrichtung (Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt GmbH) und es werden nur anonyme Ergebnislisten geliefert. Dieses Konzept soll Ende 2016 mit vier Standorten pilotiert und in 2017 sukzessiv ausgeweitet und in den Routinebetrieb überführt werden.

Im Deutschen Zentrum für Lungenforschung (DZL) werden die Bioproben dezentral an den jeweiligen Standorten gelagert und auch mit den dort jeweils etablierten Biobankmanagementsystemen verwaltet. Für Anfragen zu den bestehenden Biomaterial-Sammlungen wurde ein „Gelbe Seiten Suchkatalog“ auf der Homepage des DZL auf Basis einer Access-Datenbank angelegt. Für neu gesammelte Proben ist ein zentrales Data Warehouse zur Probensuche und Verknüpfung mit Phänotypdaten im Aufbau. Dabei sollen die Patienten zentral registriert und deren medizinische Daten pseudonymisiert im zentralen Data Warehouse (DWH) gespeichert werden. In diesem DWH (welches auf Basis von i2b2 etabliert werden soll) werden sie mittels eines Query-Tools Forschern zur Abfrage zur Verfügung gestellt. Neben Proben und Daten der Biobanken werden nach und nach sämtliche Studiendatenbanken und patientenbezogenen For-

schungsdaten des DZL in das zentrale DWH integriert und über das i2b2 Query-Tool zur Recherche freigegeben.

Auch wenn nun viele Standorte den Wechsel auf ein kommerzielles Produkt vollzogen haben, ist der Status der Nutzung dieser Systeme und der damit bereits unterstützten Arbeitsprozesse in vielen Fällen noch nicht sehr weit fortgeschritten. In der Regel wird mit der Einführung eines Biobankmanagementsystems zunächst der Prozess der Probeneinlagerung und der Lagerortverwaltung durch dialogorientierte Erfassungsmasken unterstützt. Schnittstellen zu Liquid-Handling-Plattformen, mobilen Erfassungsgeräten und Lager-Robotern wurden bisher nur an einigen Standorten etabliert. Auch scheinen die datenschutzrechtlichen Fragestellungen an der Schnittstelle zwischen Versorgung und Forschung noch nicht an allen Universitätskliniken mit ihren Biobanken vollständig geklärt zu sein.<sup>7</sup>

In Hinblick auf eine Deutschland- bzw. EU-weite Vernetzung der Biobanken ist von Interesse, dass heute schon die deutschen Zentren der Gesundheitsforschung in ihrem jeweiligen Umfeld vor exakt derselben Herausforderung stehen, wie sie zukünftig in Deutschland vom German Biobank Node und europaweit mit Hilfe des BBMRI Common Service IT zu bewältigen ist. Es gilt, die Probanden sowie die zugehörigen klinischen Informationen Standort-übergreifend zumindest virtuell zusammenzuführen und – beispielsweise zur Kohortenidentifikation – abfragbar zu machen.

Diesbezüglich sind die Konzepte und Arbeiten des Deutschen Konsortiums für Translationale Krebsforschung (DKTK) und des Deutschen Zentrums für Lungenforschung (DZL) am weitesten fortgeschritten [36] und können deshalb beispielhaft für weitere deutschlandweite Vernetzungspilotierungen in Betracht gezogen werden. Da an einigen deutschen Biobank-Standorten i2b2 bereits als Data Warehouse für die klinische Annotation von Spenderdaten etabliert ist, wird aber auch ein Konzept zur Einbindung von „Brückenköpfen“ benötigt, die keine „Light-Version“ von CentraXX, sondern ein anderes System als lokales DWH (z. B. i2b2) einsetzen. Mit dem Shared Health Research Information Network (SHRINE) ist eine Lösung zur föderierten Suche in i2b2-Instanzen seit Jahren als Open Source verfügbar [38] und wird in den USA bereits in regionalen und nationalen Netzwerken erfolgreich eingesetzt [39].

Unter Berücksichtigung dieser föderierten Abfrage- bzw. verteilten Suchkonzepte wurde im Sommer 2016 ein Pilotprojekt unter Beteiligung der Biobank des UCT Frankfurt, der Erlanger Biobank, der ZeBanC (Charité) und der Biobank des DZL initiiert, in

<sup>7</sup> Zu den komplexen rechtlichen Rahmenbedingungen der Sekundärnutzung klinischer Daten siehe auch das entsprechende Kapitel zur Sekundärnutzung im vorliegenden Report.

dem eine Vernetzung dieser vier Biobanken mit zwei unterschiedlichen lokalen Data-Warehouse-Systemen (zweimal Brückenköpfe auf Basis von „CentraXX-Light“ und zweimal i2b2-Instanzen) vorgesehen ist. Zu diesem Zweck sind sowohl die Ontologien oder Meta-Data-Dictionary-Einträge als auch die jeweiligen Informationsmodelle beider Systemwelten aufeinander abzubilden. Aus den Erfahrungen dieses Pilotprojekts soll dann das weitergehende Konzept zur Biobankenvernetzung in der German Biobank Alliance abgeleitet werden.

Interessant ist, dass die im Rahmen des amerikanischen caBIG-Projekts (Cancer Biomedical Informatics Grid) entwickelte Open Source Software caTissue, deren Installation im Rahmen des 2009 durchgeführten Evaluationsprojekts zur Weiterentwicklung der IT-Strategie der TMF [40] noch scheiterte, offensichtlich in den letzten Jahren eine starke Weiterentwicklung erfahren hat. In den USA wird sie mittlerweile unter dem Dach einer Open-Source-Initiative unter dem Namen OpenSpecimen weiterentwickelt und verbreitet. Derzeit wird sie in über 25 Biobanken eingesetzt [41]. Vor dem Hintergrund, dass diese Software vor einem Einsatz im deutschen Umfeld allerdings zunächst erneut hinsichtlich Funktionalität, gesicherter Weiterentwicklung und langfristigem Support evaluiert werden müsste, kann sie nicht als kurzfristiger Ersatz für die bestehenden kommerziellen Lösungen angesehen werden. Je nachdem wie die Erfahrungen mit dem längerfristigen Einsatz der aktuell etablierten kommerziellen Lösungen ausfallen, könnte diese Open-Source-Software jedoch perspektivisch auch wieder eine interessante Alternative und Marktbericherung darstellen.

## Bewertung und Handlungsbedarf

In einer 2012 durchgeführten internationalen Befragung [33] waren mehr als ein Drittel (38%) der Biobanken mit ihrem damaligen Biobankmanagementsystem entweder unzufrieden oder sogar sehr unzufrieden. Die Ergebnisse der internationalen Befragung bestätigen auch die bisherigen deutschen Erfahrungen, dass die Einführung eines Biobankmanagementsystems kein Projekt ist, welches in wenigen Monaten vollständig umgesetzt wird. Vielmehr dauerte der Einführungsprozess bei mehr als 50% der befragten Biobanken ein Jahr oder sogar länger. Für die vollständige Einbindung einzelner kleiner, dezentraler Biobanken an Universitätsklinikums-Standorten sind zum Teil 2 bis 3 Jahre veranschlagt.

Zukünftig wird es einen Bedarf an komplexen Anpassungen und Erweiterungen in Bezug auf die Anbindung und Ansteuerung von automatisierten Systemen (Aliquotierroboter, automatisierte Lager etc.) geben. Deswegen sollten zu den mehrfach genutzten Softwaresystemen Nutzergruppen gebildet werden, um durch eine frühzeitige

gegenseitige Abstimmung von Anforderungen und Anpassungen einen größeren Benefit für die Community zu erzielen. Bereits 2011 wurde federführend von Erlangen und Göttingen eine Starlims-Usergroup ins Leben gerufen, die sich halbjährlich trifft. Die CentraXX User Group wird von der Firma organisiert und hat sich bis 2015 jährlich getroffen. Seit 2016 finden Treffen halbjährlich statt.

An anderen Biobankstandorten scheint es aber immer noch abgekapselte Aktivitäten zur Einführung kommerzieller Softwareprodukte und deren Parametrierung zur optimalen Unterstützung der lokalen Arbeitsprozesse zu geben. Obwohl konkrete Umsetzungen jeweils in direktem Austausch mit den jeweiligen Softwareentwicklern diskutiert werden müssen, sehen wir weiterhin den Bedarf für einen stärkeren IT-System-unabhängigen Erfahrungsaustausch der deutschen Biobanken. Dies soll ab 2017 unter anderem im Rahmen der German Biobank Alliance (GBA) adressiert werden.

Die Analyse aller mit dem Biobanking verbundenen Arbeitsprozesse befindet sich im Vergleich zu den in der Krankenversorgung durch IT-Lösungen unterstützten Prozessen noch in einer frühen Phase. Erst wenn diese Prozesse umfassend analysiert und definiert wurden, ist eine optimale IT-Unterstützung aller Teilprozesse des Biobanking möglich. Als derzeit noch fehlende Bausteine einer umfassenden IT-Infrastruktur für Biobanken wurden in der GBA ein erweitertes und föderiertes Meta Data Repository, ein Modul zum Biomaterial-Projektmanagement, ein Modul zum ID-Management und zum Tracking von Proben bzw. damit durchgeführten Analysen, sowie ein Modul zum Consent- und Patientenkontaktmanagement identifiziert. Entsprechende Lösungen hierfür sollen in den Jahren 2017 bis 2019 im Rahmen der GBA entwickelt und für die deutschen Biobanken als Open-Source-Lösungen verfügbar gemacht werden.

Darüber hinaus wurden Biobankmanagementsysteme an vielen Standorten im ersten Schritt noch als isolierte Stand-Alone-Systeme etabliert; die Konzeption und Umsetzung von Schnittstellen sowohl zu IT-Systemen der Krankenversorgung als auch z. B. zu Robotersystemen steht noch aus.

Des Weiteren scheint die hoheitliche Trennung von Phänotypdaten und Probanden derzeit noch nicht überall konsequent umgesetzt zu sein. Bei der bereits erwähnten Umfrage zur Biomaterialverwaltungssoftware im Oktober 2014 in der AG Biomaterialbanken der TMF und dem Deutschen Biobankenregister gaben zehn von 30 Biobanken an, dass sich ebenfalls klinische Informationen in der Biomaterialverwaltung befinden würden. Dies könnte daran liegen, dass professionelle kommerzielle Softwareprodukte zur Biomaterialverwaltung nicht von Haus aus Schnittstellen zu existierenden

ID-Management-Tools<sup>8</sup> anbieten. Je nach Umfeld der Biobank, ist dies aber ggf. an der Quelle auch nicht notwendig. Befindet sich die Biobank in einem Versorgungskontext und das Personal ist zur Geheimhaltung verpflichtet, so müssen die Proben erst bei der Herausgabe an Dritte pseudonymisiert werden. Die effiziente Umsetzung der hohen Datenschutzerfordernungen, wie z. B. die in bestimmten Fällen erforderliche zusätzliche Umschlüsselung der Proben-Identifikation in eine LabID<sub>tr</sub> gemäß TMF-Datenschutzkonzept [4], stellt viele Biobanken ebenfalls noch vor große Herausforderungen.

Standards, wie die zuvor erwähnten SPREC, MIABIS und BRISQ sind leider noch relativ unbekannt in der Biobanken-Community (Ergebnisse der Interactive Session der ESBBperanto und Enviro-Bio Arbeitsgruppen der ESBB auf der ESBB Tagung 2013 in Verona [42]). Hier sollten Weiterbildungsmaßnahmen ergriffen werden, z. B. in der Form von Workshops, um die Community auf die Existenz von Standards hinzuweisen und die Vorteile von deren Verwendung aufzuzeigen.

Einen regelmäßigen Erfahrungsaustausch innerhalb der GBA, aber auch über diese Biobanken hinausgehend, halten wir weiterhin für unbedingt notwendig, um zu vermeiden, dass viele Standorte ihre Erfahrungen in eher unkoordinierter Form nebeneinanderher sammeln. Unterstützt durch die Deutsche Gesellschaft für Medizinische Informatik, Epidemiologie und Biometrie (GMDS), die TMF und den German Biobank Node wurde dieser Austausch Ende 2014 begonnen<sup>9</sup> und wird in dieser Form auch in 2017 fortgesetzt werden. Neben der Durchführung expliziter Workshops bieten sich für diesen Austausch auch die regelmäßigen Sitzungen der AG Biomaterialbanken der TMF an, die seit 2008 auch für Teilnehmer offen sind, die keinem TMF-Mitgliedsverbund angehören.<sup>10</sup>

In den kommenden Jahren werden sich voraussichtlich zwei Fördermaßnahmen des BMBF auf die weitere Entwicklung der Biobanken-IT in Deutschland nachhaltig auswirken: Die in die GBA mündende „Ertüchtigungs-Initiative“ und das Förderkonzept Medizininformatik. Damit keine unabgestimmten und schlimmstenfalls inkompatiblen IT-Landschaften entstehen, benötigen beide Fördermaßnahmen übergreifend definierte Abstimmungsprozesse, damit die entsprechende Koordinierung der Entwicklungen sichergestellt werden kann.

---

8 Siehe hierzu das Kapitel zum ID-Management im vorliegenden Report.

9 Vgl. <http://www.pg-ss.imi.uni-erlangen.de/Seiten/11Sitzung.aspx> und [http://bbmri.de/workshop\\_it](http://bbmri.de/workshop_it).

10 [http://www.tmf-ev.de/Arbeitsgruppen\\_Foren/AGBMB.aspx](http://www.tmf-ev.de/Arbeitsgruppen_Foren/AGBMB.aspx)

## Literatur

1. Riegman, P.H.J., Morente, M.M., Betsou, F., de Blasio, P., Geary, P., *Biobanking for better healthcare*. *Molecular oncology*, 2008. **2**(3): S. 213–222.
2. Demiroglu, S.Y., Skrowny, D., Quade, M. et al., *Managing sensitive phenotypic data and biomaterial in large-scale collaborative psychiatric genetic research projects: practical considerations*. *Molecular psychiatry*, 2012. **17**(12): S. 1180–1185.
3. Prokosch, H.U., Beck, A., Ganslandt, T. et al., *IT Infrastructure Components for Biobanking*. *Applied Clinical Informatics*, 2010. **1**(4): S. 419–429.
4. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten – Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
5. Betsou, F., Lehmann, S., Ashton, G. et al., *Standard preanalytical coding for biospecimens: defining the sample PREanalytical code*. *Cancer epidemiology, biomarkers & prevention*, 2010. **19**(4): S. 1004–1011.
6. Lehmann, S., Guadagni, F., Moore, H. et al., *Standard Preanalytical Coding for Biospecimens: Review and Implementation of the Sample PREanalytical Code (SPREC)*. *Biopreservation and Biobanking*, 2012. **10**(4): S. 366–374.
7. Nussbeck, S.Y., Benson, E.E., Betsou, F., Guadagni, F., Lehmann, S., Umbach, N., *Is there a protocol for using the SPREC?* *Biopreserv Biobank*, 2013. **11**(5): S. 260–266.
8. Skrowny, D., Grütz, R., Nussbeck, S.Y., *Anforderungen an und Übersicht über aktuelle Software zur Verwaltung von Biomaterialien*, in *Biobanken-Forschung in Deutschland: Vom Konzept zur Realisierung. 3. Nationales Biobanken-Symposium 2014 – Tagungsband* Hrsg.: M. Hummel, et al. 2014, AKA, Berlin. S. 59–68.
9. Kiehntopf, M., Böer, K., *Biomaterialbanken – Checkliste zur Qualitätssicherung*. 2008, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
10. Pommerening, K., *Das Datenschutzkonzept der TMF für Biomaterialbanken (The TMF Data Protection Scheme for Biobanks)*. *it – Information Technology*, 2007. **49**(6): S. 352–359.
11. Simon, J.W., Paslack, R., Robiński, J., Goebel, J.W., Krawczak, M., *Biomaterialbanken – Rechtliche Rahmenbedingungen*. 2006, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
12. Mayrhofer, M.T., Holub, P., Wutte, A., Litton, J.-E., *BBMRI-ERIC: the novel gateway to biobanks*. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 2016: S. 1–6.
13. Hewitt, R., Watson, P., *Defining biobank*. *Biopreserv Biobank*, 2013. **11**(5): S. 309–315.
14. Helbing, K., Demiroglu, S.Y., Rakebrandt, F., Pommerening, K., Rienhoff, O., Sax, U., *A data protection scheme for medical research networks. Review after five years of operation*. *Methods of information in medicine*, 2010. **49**(6): S. 601–607.

15. Kuhn, K.A., Bild, R., Anton, G., Schuffenhauer, S., Wichmann, H.-E., *Vernetzung von Biobanken großer europäischer Kohorten (EU-Projekt BBMRI-LPC)*. Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz, 2016: S. 1–5.
16. Becker, R., Ihle, P., Pommerening, K., Harnischmacher, U. *Ein generisches Datenschutzkonzept für Biomaterialbanken (Version 1.0)*. 2006. TMF, <http://www.tmf-ev.de/produkte/P010021> (Abruf: 2016-11-27).
17. Ethikrat *Humanbiobanken für die Forschung*. 2010. Deutscher Ethikrat, <http://www.ethikrat.org/dateien/pdf/stellungnahme-humanbiobanken-fuer-die-forschung.pdf> (Abruf: 2016-11-30).
18. Campbell, L.D., Betsou, F., Garcia, D.L. et al., *2012 Best Practices for Repositories: Collection, Storage, Retrieval, and Distribution of Biological Materials for Research*. Biopreservation and Biobanking, 2012. **10**(2): S. 80–161.
19. NCI *NCI Best Practices for Biospecimen Resources*. 2011. National Cancer Institute, <http://biospecimens.cancer.gov/bestpractices/2011-NCIBestPractices.pdf> (Abruf: 2016-11-30).
20. OECD *OECD Best Practice Guidelines for Biological Resource Centres*. 2007. The Organisation for Economic Co-operation and Development, <http://www.oecd.org/sti/biotechnology/policies/38777417.pdf> (Abruf: 2016-11-30).
21. OECD *OECD Guidelines on Human Biobanks and Genetic Research Databases*. 2009. The Organisation for Economic Co-operation and Development, <http://www.oecd.org/sti/biotech/44054609.pdf> (Abruf: 2016-11-30).
22. COE *Recommendation Rec(2006)4 of the Committee of Ministers to member states on Research on Biological Materials of Human Origin*. 2006. Council of Europe–Committee of Ministers, <https://wcd.coe.int/ViewDoc.jsp?id=977859> (Abruf: 2016-11-30).
23. WMA *World medical association declaration of helsinki: Ethical principles for medical research involving human subjects*. 2013. World Medical Association, 20, <http://dx.doi.org/10.1001/jama.2013.281053> (Abruf: 2016-11-30).
24. WMA *WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks*. 2016. World Medical Association, <http://www.wma.net/en/30publications/10policies/d1/index.html> (Abruf: 2016-11-18).
25. Schmidt am Busch, B., Gassner, U.M., Wollenschläger, F., *Der Augsburg-Münchner-Entwurf eines Biobankgesetzes*. Datenschutz und Datensicherheit, 2016. **2016**(6): S. 365-370.
26. TMF *Kurzstellungnahme der AG Biomaterialbanken der TMF zur Stellungnahme „Humanbiobanken für die Forschung“ des Deutschen Ethikrats vom 15. Juni 2010*. 2010. TMF–Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V., <http://www.tmf-ev.de/News/696> (Abruf: 2016-11-30).

27. Moore, H.M., Kelly, A.B., Jewell, S.D. et al., *Biospecimen reporting for improved study quality (BRISQ)*. Cancer Cytopathol, 2011. **119**(2): S. 92-101.
28. Norlin, L., Fransson, M.N., Eriksson, M., Merino-Martinez, R., Anderberg, M., Kurtovic, S., Litton, J.E., *A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS*. Biopreserv Biobank, 2012. **10**(4): S. 343-348.
29. Merino-Martinez, R., Norlin, L., van Enkevort, D. et al., *Toward Global Biobank Integration by Implementation of the Minimum Information About Biobank Data Sharing (MIABIS 2.0 Core)*. Biopreservation and Biobanking, 2016. **14**(4): S. 298-306.
30. Benson, E.E., Betsou, F., Amaral, R., Santos, L.M.A., Harding, K., *Standard PReanalytical Codes: A New Paradigm for Environmental Biobanking Sectors Explored in Algal Culture Collections*. Biopreservation and Biobanking, 2011. **9**(4): S. 399-410.
31. Droege, G., Barker, K., Seberg, O. et al., *The Global Genome Biodiversity Network (GGBN) Data Standard specification*. Database, 2016. **2016**.
32. Furuta, K., Schacter, B., *Report on Status of ISO276/WG2 on Biobanks and Bioresources: International Standards for Biobanking*. Biopreservation and Biobanking, 2015. **13**(6): S. 452-453.
33. Prokosch, H.-U., Mate, S., Christoph, J. et al., *Designing and implementing a biobanking IT framework for multiple research scenarios*. Stud Health Technol Inform, 2012. **180**: S. 559–563.
34. Fearn, P., Michels, C., Meagher, K., Cada, M., *2012 International Society for Biological and Environmental Repositories Informatics Working Group: Survey Results and Conclusions*. Biopreservation and Biobanking, 2013. **11**(1): S. 64–66.
35. Beck, A., Prokosch, H.U. *TMF-Projekt „IT-Strategie“, Teilprojekt 3: Analyse der derzeit in Deutschland verfügbaren IT-Werkzeuge zur Unterstützung des Managements von Biomaterialbanken*. 2010. TMF, [https://www.tmf-ev.de/Themen/Projekte/V054\\_01\\_IT\\_Strategie.aspx](https://www.tmf-ev.de/Themen/Projekte/V054_01_IT_Strategie.aspx) (Abruf: 2016-11-30).
36. Lablans, M., Kadioglu, D., Muscholl, M., Ückert, F., *Exploiting Distributed, Heterogeneous and Sensitive Data Stocks while Maintaining the Owner's Data Sovereignty*. Methods of Information in Medicine, 2015. **54**(4): S. 346-352.
37. Lablans, M., Kadioglu, D., Mate, S., Leb, I., Prokosch, H.-U., Ückert, F., *Strategien zur Vernetzung von Biobanken. Klassifizierung verschiedener Ansätze zur Probensuche und Ausblick auf die Zukunft in der BBMRI-ERIC*. Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz, 2016: S. 1-6.
38. Weber, G.M., Murphy, S.N., McMurry, A.J., Macfadden, D., Nigrin, D.J., Churchill, S., Kohane, I.S., *The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories*. J Am Med Inform Assoc, 2009. **16**(5): S. 624-630.

39. McMurry, A.J., Murphy, S.N., MacFadden, D. et al., *SHRINE: enabling nationally scalable multi-site disease studies*. PLoS One, 2013. **8**(3): S. e55811.
40. Sax, U., Rakebrandt, F., Grütz, R., Thormann, S. *TMF-Projekt „IT-Strategie“, Teilprojekt 2b: Analyse ausgewählter caBIG Tools*. 2009. TMF, [https://www.tmf-ev.de/Themen/Projekte/V054\\_01\\_IT\\_Strategie.aspx](https://www.tmf-ev.de/Themen/Projekte/V054_01_IT_Strategie.aspx) (Abruf: 2016-11-30).
41. McIntosh, L.D., Sharma, M.K., Mulvihill, D. et al., *caTissue suite to OpenSpecimen: Developing an extensible, open source, web-based biobanking management system*. J Biomed Inform, 2015.
42. Nussbeck, S.Y., Rabone, M., Benson, E.E., Droege, G., Mackenzie-Dodds, J., Lawlor, R.T., *Life in Data'-Outcome of a Multi-Disciplinary, Interactive Biobanking Conference Session on Sample Data*. Biopreservation and Biobanking, 2016.



# Molekularbiologische Daten aus Hochdurchsatz-Analysen

Nadine Umbach, Tim Reißbarth, Ulrich Sax

Mitarbeit: Andre Franke<sup>1</sup>, Jürgen Eils, Chris Lawerenz<sup>2</sup>, Christoph Schickhardt<sup>3</sup>

## Anwendungsbereich

Molekularbiologische Methoden und Hochdurchsatz-Technologien haben bei der qualitativen und quantitativen Analyse von biologischen Proben in den letzten Jahren eine rasante Entwicklung erlebt. So ist es heute möglich, Biomoleküle einer Zelle mit vertretbarem zeitlichen und finanziellen Aufwand umfassend zu charakterisieren. Dabei haben sich verschiedene Subdisziplinen herausgebildet, die jeweils mit der Silbe „-omics“ enden und sich in mindestens fünf Bereiche untergliedern lassen. So befasst sich die Disziplin *Genomics* mit der Erforschung von Genen bis hin zur Analyse des gesamten Erbguts bzw. Genoms eines Lebewesens. Bei *Epigenomics* werden Modifikationen der DNA oder von Histonen (Methylierung, Acetylierung), welche zur Veränderung der Genomfunktion führen, untersucht. Im Bereich *Transcriptomics* wird die quantitative Ausprägung von Genen in Form von Transkripten (mRNAs), d. h. die Expressionsaktivität von Genen erfasst. Die Erforschung des Proteoms, also der Gesamtheit der Proteine, die durch das Genom beschrieben werden, sowie deren Funktionen und Wechselwirkungen ist das Ziel der Disziplin *Proteomics*. Im Rahmen von *Metabolomics* werden schließlich kleine chemische Moleküle, Metabolite, also Stoffwechselprodukte, die bei der Regulierung von Stoffwechselkreisläufen oder Zellaktivitäten durch Proteine umgesetzt oder hergestellt werden, betrachtet (vgl. Abb. 7.1).

Möglich werden die Analysen durch den Einsatz von molekularen Hochdurchsatz-Technologien, die nahezu automatisch Biomoleküle untersuchen können. So werden in der Genomik unter dem Begriff **Next-Generation Sequencing (NGS)** Technologien zusammengefasst, die auf massiver Parallelisierung und Automatisierung beruhen und das humane Genom innerhalb einiger Stunden für unter 1000 US\$ entschlüsseln können. Neben dem **Whole-Genome Sequencing (WGS)**, d. h. der Entschlüsselung des

- 
- 1 Andre Franke (Christian-Albrechts-Universität zu Kiel) danken wir für die Mithilfe bei den Angaben bzgl. Datenmengen sowie für die Darstellung zum medizinischen Forschungsdatenmanagement an der Christian-Albrechts-Universität.
  - 2 Jürgen Eils und Chris Lawerenz (Universität Heidelberg und Deutsches Krebsforschungszentrum Heidelberg) danken wir für die Mithilfe bei den Angaben bzgl. Datenmengen und eingesetzten IT-Werkzeugen.
  - 3 Für die fachliche Unterstützung bei der Erhebung der ethischen und rechtlichen Aspekte im Umgang mit OMICS-Daten danken wir Dr. Christoph Schickhardt vom Nationalen Centrum für Tumorerkrankungen in Heidelberg.

gesamten Genoms) und dem **Whole-Exome Sequencing** (WES, d. h. Entschlüsselung der Protein-kodierenden Bereiche im Genom) findet insbesondere in der klinischen Diagnostik die Analyse von **Panels**, also die Sequenzierung aller für eine bestimmte Krankheit bekanntermaßen relevanten Gene, weite Verbreitung. Zur Genotypisierung von genetischen Varianten sind auch die sogenannten **Single Nucleotide Polymorphism** (SNP)-**Chips**<sup>4</sup> verbreitet im Einsatz, welche bis zu mehrere Millionen Varianten mithilfe der **Microarray-Technologie** messen. In der Epigenetik sind ChIP-Sequencing und Microarrays weit verbreitete Methoden. Für die Analyse von genomweiten Genexpressionen wird in der Transkriptomik **RNA-Sequencing** oder ebenfalls die **Microarray-Technologie** eingesetzt. Zu den Hochdurchsatz-Technologien in der Metabolomik und Proteomik gehören unter anderem flüssigkeitschromatographische und hochsensitive massenspektrometrische Verfahren (LC-MS), welche die Identifizierung und Quantifizierung von Proteinen sowie ihrer posttranslationalen Modifikationen erlauben.

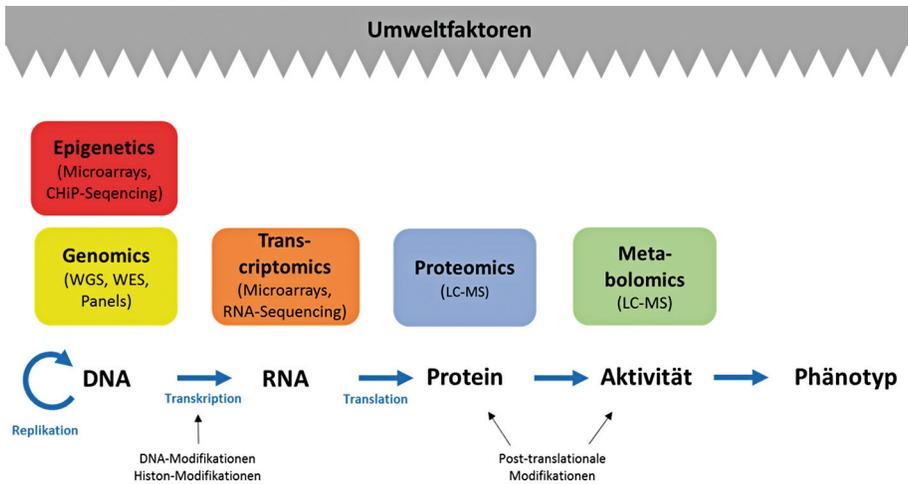


Abbildung 7.1: Darstellung des zentralen Dogmas der Biologie mit der methodischen Einbindung der Omics-Disziplinen: Auf der Hauptachse ist stark vereinfacht der biologische Verlauf von der Replikation der DNA über die Transkription in RNA, die Translation in Proteine bis hin zum klinischen Phänotypen verzeichnet. Oberhalb sind die methodisch korrespondierenden Omics-Verfahren mit Beispielen genannt: Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), Microarrays, ChIP- und RNA-Sequencing, Liquid Chromatography–Mass Spectrometry (LC-MS, auch HPLC-MS für High Performance LC-MS). Die Regulation von DNA, RNA, Proteinen und des Phänotyps wird von verschiedenen Umweltfaktoren beeinflusst.

4 Als Polymorphismus werden unterschiedliche Genvarianten innerhalb einer (vorwiegend gesunden) Population mit einer Frequenz von  $\geq 1\%$  bezeichnet.

Durch die stetige Weiterentwicklung und Verbreitung von Omics-Technologien steigt auch die Anzahl von Anwendungsmöglichkeiten. Im Forschungskontext, in dem molekulare Hochdurchsatz-Technologien bereits gut etabliert sind, zählen folgende Anwendungen dazu:

- ▶ Sequenzierung neuer Genome von bisher unbekanntem Organismen (i. d. R. Bakterien oder Viren);
- ▶ genomweite Re-Sequenzierung zur Detektion und Analyse von Rearrangements, Insertionen, Deletionen oder Variationen einzelner Basen in der DNA; Analyse von Methylierungen der DNA-Sequenz, die als vererbte Modifikationen die Genexpression regulieren, ohne die eigentliche Sequenz zu beeinflussen;
- ▶ genomweite Assoziationsstudien (GWAS, engl. Genome-wide association study) basierend auf Genotyp-Phänotyp-Beziehungen für die Identifizierung krankheitsrelevanter Genorte;
- ▶ Metagenomanalyse bei der Untersuchung von Genomfragmenten und deren Zusammensetzung in einer Organismenpopulation und deren Charakterisierung in einer Bioprobe (s. Tab. 7.2);
- ▶ Hochdurchsatzanalysen von Mikrobiomdaten (16S-Amplicon-Sequenzierung z. B. aus Stuhlproben), anhand derer die Zusammensetzung des Mikrobioms in der Probe abgeschätzt werden kann (s. Tab. 7.3).

Ziel der anwendungsbezogenen Forschung ist im Sinne der stratifizierenden oder sogenannten personalisierten Medizin, die Formulierung und Herstellung von patientengerechten Arzneimitteln, die einen maximalen Effekt mit möglichst wenig Nebenwirkungen erzielen. Ebenso spielen molekulare Hochdurchsatz-Analysen bei der Herstellung und Qualitätssicherung von Stammzellen eine wichtige Rolle. In der Biotechnologie sind sie unerlässlich für die Bewertung der Sicherheit und der Nährwertqualität von gentechnisch veränderten Lebensmitteln sowie bei der Züchtung von ergebnisreichen und widerstandsfähigen Pflanzen.

In der klinischen Versorgung steht NGS vor allem in zwei Anwendungsbereichen vor dem Durchbruch:

- ▶ in der Humangenetik für die Analyse von seltenen, monogenen Erkrankungen [1] sowie
- ▶ in der Onkologie für die eindeutige Diagnoseerstellung, die molekular-pathologische Subklassifizierung von klinisch nicht unterscheidbaren Tumortypen, und auch für die Steuerung von Therapieentscheidungen sowie die Abschätzung von Prognosen [2].

Trotz der Verbesserungen bei Qualität und Reproduzierbarkeit von Biomarkerexperimenten und -untersuchungen ist zu berücksichtigen, dass nach wie vor Fehler, d. h. falsch positive oder falsch negative Befunde, vorliegen können und dass aufgrund der hohen Anzahl an gemessenen Biomarkern selbst bei einem kleinen Fehler erster bzw. zweiter Art der prognostische Wert von einem einzelnen Biomarker sehr gering ist [3-5].

## Betreiber, Anwender, Nutzer

Der Betrieb von Omics-Technologien ist derzeit sehr heterogen organisiert. Im Forschungskontext werden molekulare Hochdurchsatz-Analysen üblicherweise in zentralen Service-Einrichtungen von wissenschaftlichen Forschungseinrichtungen oder von Institutionen des Gesundheitswesens (hier insbesondere Universitätsklinika) angeboten. Diese sind in der Regel auf ausgewählte Fach-Disziplinen (z. B. Genomik oder Proteomik) fokussiert. Im Rahmen der Helmholtz-Gemeinschaft und der Max-Planck-Gesellschaft sind entsprechende Service-Einrichtungen nur an ausgewählten Standorten angesiedelt. Diese bieten ihre Dienstleistungen standortübergreifend an. Auch in größeren Verbundforschungsprojekten wie z. B. den deutschen Zentren für Gesundheitsforschung und der NAKO Gesundheitsstudie sind in vielen Fällen bereits bestehende und etablierte zentrale Service-Einrichtungen eingebunden. Diese Organisationsformen ermöglichen eine effiziente und effektive Bereitstellung sowie den nachhaltigen Betrieb von spezialisierten Geräten und eine hohe Methodenkompetenz. Zusätzlich zum Laborservice werden eingehende Beratungen zum Projektdesign, die (Weiter-)Entwicklung und unmittelbare Anwendung von neuen Technologien sowie Schulungen und Hands-On-Kurse für Nutzer zwecks Transfer von fachspezifischem Know-how angeboten. Für das meist interdisziplinäre Projekt-Design und die Auswertung der komplexen Daten sind auch Methoden und ein hohes Maß an Expertise aus der Biologie und ggf. der Klinik sowie der Bioinformatik und der Biostatistik notwendig. Die letzteren beiden Dienstleistungen können häufig nicht alleine von den Einrichtungen übernommen werden, in denen die Omics-Daten generiert wurden, sondern bedürfen einer ausreichend ausgestatteten Bioinformatik-Infrastruktur. Vor allem eine enge Vernetzung der Bioinformatik mit den Daten-produzierenden Einheiten ist essentiell, insbesondere bei dem Finden und Beseitigen von Problemen in der über mehrere Stationen reichenden Prozesskette.

Die Leistungen der zentralen Service-Einrichtungen können dabei nicht nur von den Institutionen genutzt werden, an die sie direkt angebunden sind. Externe können das Angebot üblicherweise auf Vollkosten-Basis nutzen. Hierbei gelten typischerweise die in einer Nutzerordnung und einem Betriebskonzept fixierten Rahmenbedingungen.

Oftmals werden Omics-Technologien auch außerhalb von zentralen wissenschaftlichen Service-Einrichtungen angeboten. Naturwissenschaftliche oder biomedizinische Abteilungen, deren Schwerpunkt auf molekularen Hochdurchsatz-Analysen liegt, verfügen oft selbst über entsprechende Geräte und fundiertes Wissen über die Methodik und die Analyse der komplexen Daten. Auch im klinischen Kontext nutzen immer mehr Abteilungen (z. B. Humangenetik, Pathologie, Onkologie) diese Technologien, insbesondere für die Analyse von Panels. Aufgrund der besonderen Anforderungen in der klinischen Diagnostik im Hinblick auf Qualität, Validität und Priorisierung der Analysen müssen ggf. eigene Infrastrukturen im Versorgungskontext vorgehalten werden. Darüber hinaus treten im wissenschaftlichen als auch im Diagnostik-bezogenen Kontext zahlreiche Firmen in Erscheinung, die auf Auftragsbasis zahlreiche Analysen in den Disziplinen Genomik, Epigenetik und Transkriptomik durchführen.

Die Nutzer im Forschungskontext sind in erster Linie biomedizinische Grundlagenforscher und klinisch forschende Ärzte, die die Daten zur Beantwortung ihrer Fragestellungen in wissenschaftlichen Projekten oder Studien benötigen. Aktuell wird z. B. im Deutschen Zentrum für Herz-Kreislauf-Forschung e. V. (DZHK) eine „OMICS-Resource“ bestehend aus 1.000 vollständigen Genomen und 200 kompletten RNA-Sequenzen aufgebaut mit dem Ziel, Assoziationen von genomischen Varianten auf herzbezogene Phänotypen zu untersuchen, die in bevölkerungsbasierten Studien ermittelt wurden. Im Versorgungskontext benutzen Ärzte unter anderem NGS-Technologien zur Erstellung von Diagnosen, Prognosen und Therapieentscheidungen. Zunehmend fragen auch Patienten Biomarker-Untersuchungen an, was hinsichtlich der vorhandenen Qualitätsstandards und der nicht immer einfachen Interpretation der Daten kritisch zu betrachten ist und neue ethische und rechtliche Fragestellungen aufwirft [6-8].

## Rahmenbedingungen

### **Organisatorische und technische Rahmenbedingungen**

Organisatorische und technische Aspekte müssen beim Einsatz von Omics-Technologien für Forschungszwecke auf verschiedenen Ebenen berücksichtigt werden.

Die molekulare Hochdurchsatz-Analyse in zentralen Service-Einheiten und in naturwissenschaftlichen oder biomedizinischen Abteilungen, die Omics-Technologien zur Beantwortung ihrer Forschungsfragen selbst vorhalten, beginnt mit einer Anfrage des potentiellen Auftraggebers an die Einrichtung. Bei zentralen Service-Einrichtungen werden Informationen zu Art, Umfang und Rahmenbedingungen der Dienstleistung

gen (u. a. vorhandene Omics-Plattformen, Gebührenmodelle, Probenaufbereitung und –bereitstellung) in der Regel über eigene Homepages öffentlich bereitgestellt. Im Gegensatz zu Service-Einrichtungen ist eine solche Transparenz bezüglich Expertise und Geräteausstattung bei Abteilungen aus dem Forschungs- und Versorgungskontext häufig nicht anzutreffen.

Für die Durchführung von molekularen Hochdurchsatz-Analysen müssen die flüssigen, festen bzw. zellulären Bioproben zunächst aufbereitet werden. Dieser Schritt wird – abhängig von den einzelnen Omics-Disziplinen – in den Abteilungen der Auftraggeber oder in den zentralen Service-Einrichtungen durchgeführt. Entammt das Biomaterial für Analysen im Genomik- bzw. Transkriptomik-Bereich einer zentralen Biobank, kann die Isolation von DNA- bzw. RNA-Material auch von den Mitarbeitern der Biobank übernommen werden.

Die technischen Entwicklungen im Rahmen von Omics-Technologien führen auch zu großen Herausforderungen im Umgang mit den Daten. Eine Übersicht zu Datenformaten und Datenmengen bei WGS-Analysen ist Tabelle 7.1 zu entnehmen.

Beispielhaft sind ergänzend Datenformate und Datenmengen für Mikrobiomanalysen in Tabelle 7.3 aufgeführt. Es ist erkennbar, dass die ausgewerteten Daten im Vergleich mit den Rohsequenzdaten eine erheblich reduzierte Datenmenge aufweisen. Allerdings wird zugleich deutlich, welche zusätzlichen Qualitätskontroll- und Metadaten mitgeführt werden müssen.

Daten aus molekularbiologischen Hochdurchsatz-Analysen stellen hinsichtlich Vertraulichkeit, Speichervolumen, Zugriff und Auswertung besondere Anforderungen an die Infrastruktur, die bei vielen dezentralen Ansätzen derzeit nicht erfüllt werden können. Bei der zugrundeliegenden Architektur kann zwischen Kurzzeit- und Langzeitspeicher unterschieden werden. Während für die Kurzzeitspeicherung über wenige Monate lokale Server genutzt werden, erfolgt die Aufbewahrung der Daten gemäß guter wissenschaftlicher Praxis üblicherweise für zehn Jahre z. B. auf Bändern in Rechenzentren. Neben ausreichend Speicherkapazität – größere Einrichtungen in Deutschland bieten hierfür derzeit 12–20 PB gespiegelt über zwei Standorte an – muss auch die entsprechende Netzwerkbandbreite für die Übertragung der Daten vom Entstehungsort zur Speicherung und von der Speicherung zur Auswertung vorhanden sein. Darüber hinaus sind Ressourcen für die Verarbeitung und Analyse der Daten zu beachten. Selbst für größere wissenschaftliche Einrichtungen und Universitätsklinika ist die Bereitstellung von entsprechenden Infrastruktur-Lösungen nicht trivial und erfordert technisches, konzeptionelles, regulatorisches und organisatorisches Know-How. In den meisten Fällen

sind die gefundenen Lösungen eher als pragmatisch-funktional anzusehen und nicht nach umfassenden strategischen Vorgaben ausgerichtet.

**Tabelle 7.1: Datenformate und –mengen bei WGS-Analysen bezogen auf ein WGS mit 30-facher Abdeckung**

Workflow	Datenformat	Rechenzeit <sup>1</sup>	Datenmenge <sup>2</sup>	Kommentar
Rohsequenzdaten mit Qualitätsparametern	FASTQ	½-1d/run	ZIP komprimiert ca. 100 GB/WGS;	Textbasiertes Format zur Speicherung sowohl von Nukleotidsequenzen als auch der korrespondierenden Qualitätsscores. Dabei wird sowohl das Nukleotid als auch der Qualitätsscore (Phred-like score) in einem einzelnen ASCII Zeichen untergebracht [9]. De-facto Standard zur Speicherung der Ausgabe von Hochdurchsatzsequenzierern [10]. Eine Konvertierung von BCL in FASTQ.gz kann durch Programme wie Illumina bcl2fastq erfolgen.
Mapping auf Referenzsequenz	BAM (Binary Alignment / Map)	2h-20h pro WGS je nach Software und Hardware	komprimiert ca. 100 GB/WGS,	Generisches textbasiertes Format für die Dokumentation spezifischer Alignments (Übereinstimmungen) verschiedener Sequenzierungsplattformen im Vergleich mit einer Referenzsequenz [11].
Verzeichnen der Varianten	Variant Call Format (VCF)	10h-20h pro Genom je nach Varianten und Software und Hardware	ca. <1-5 GB pro Genom	Generisches textbasiertes Format zur Speicherung von DNA-Varianten (SNVs, Indels, Strukturvarianten). [12].

<sup>1</sup> d = dies, Tag; h = hora, Stunde

<sup>2</sup> GB = Gigabyte

Weitere Erklärungen finden sich im Text und im zusammengefassten Abkürzungsverzeichnis am Ende des Buchs.

**Tabelle 7.2: Datenformate und -mengen bei Shotgun-Metagenomics-Analysen bezogen auf ein Metagenom mit 20 Mio. Paired-End Reads (Rohdaten)**

Workflow	Datenformat	Rechenzeit <sup>1</sup>	Datenmenge <sup>2</sup>	Kommentar
Rohsequenzdaten mit Qualitätsparametern	FASTQ	NA	GZIP komprimiert ca. 4 GB/Metagenom; unkomprimiert ca 10-12 GB/Metagenom	Textbasiertes Format zur Speicherung sowohl von Nukleotidsequenzen als auch der korrespondierenden Qualitätsscores. Dabei wird sowohl das Nukleotid als auch der Qualitätsscore (Phred-like score) in einem einzelnen ASCII Zeichen untergebracht [9]. De-facto Standard zur Speicherung der Ausgabe von Hochdurchsatzsequenzierern [10]. Eine Konvertierung von BCL in FASTQ.gz kann durch Programme wie Illumina bcl2fastq erfolgen.
Qualitätskontrollierte Sequenzdaten mit Qualitätsparametern	FASTQ	Abhängig von Software und Hardware; insb. von Parallelisierungsmöglichkeiten	komprimiert ca. 3 GB/Metagenom,	Filtern der Sequenzen beinhaltet das Entfernen möglicher Adaptersequenzen (Trimmomatic), Sequenzqualitäts-basiertes Trimming der Reads (Sickle) und das Entfernen von Wirtsequenzen (DeconSeq)
Assembly der metagenomischen Contigs	FASTA	Abhängig von Software und Hardware	ca. 0,1 – 1 GB pro Metagenom	Intermediäre Dateien können mehrere Gigabyte groß sein. Mögliche Software zum Assemblieren speziell von Metagenomen: MetaVelvet, Spades (v. 1.9). Ausgabe ist ein textbasiertes Format der Nukleotidsequenzen.

<sup>1</sup> d = dies, Tag; h = hora, Stunde

<sup>2</sup> GB = Gigabyte

Weitere Erklärungen finden sich im Text und im zusammengefassten Abkürzungsverzeichnis am Ende des Buchs.

**Tabelle 7.3: Datenformate und -mengen bei 16S-Amplicon-Sequenzierung (Mikrobiomdaten) und ca. 20 Megabasen Sequenzdaten pro Sample<sup>1</sup>**

Workflow	Datenformat	Rechenzeit <sup>2</sup>	Datenmenge <sup>3</sup>	Kommentar
Rohsequenz- daten mit Qualitätspara- metern	FASTQ	NA	GZIP kompri- miert ca. 14 MB/Probe;	Textbasiertes Format zur Speicherung so- wohl von Nukleotidsequenzen als auch der korrespondierenden Qualitätsscores. Da- bei wird sowohl das Nukleotid als auch der Qualitäts-Score (Phred-like score) in einem einzelnen ASCII Zeichen untergebracht [9]. De-facto Standard zur Speicherung der Aus- gabe von Hochdurchsatzsequenzierern [10]. Eine Konvertierung von BCL in FASTQ.gz kann durch Programme wie Illumina bcl- 2fastq erfolgen.
Qualitätskon- trollierte Se- quenzdaten	FASTA	5-10 Minuten pro Probe;  Abhängig von Software und Hardware; insb. von Par- allelisierungs- möglichkeiten	10 MB/Probe	Ausgabe ist ein textbasiertes Format der Nukleotidsequenzen. Qualitätskontrolle der Sequenzen beinhaltet Sequenzquali- täts-basiertes Trimming der Reads (Sickle) und anschließendes Filtern von Sequen- zen mit schlechter Durchschnittsqualität (FastX-Toolbox). Desweiteren werden PCR-Artefakte, sogenannte Chimären, ent- fernt (Usearch/Vsearch).
Taxonomi- sche Anno- tation der Sequenzen	Text	1-10 Minuten pro Probe;  Abhängig von Software und Hardware	ca. 6MB/ Probe	Einfache Textdatei mit Informationen zum Sequenznamen, taxonomischer Klassifi- zierung mit Wahrscheinlichkeiten für die Zuordnung, sowie der bestmöglichen taxo- nomischen Annotation. Mögliche Software: RDP Classifier, Usearch (v. 8)/UTAX.

<sup>1</sup> entspricht der durchschnittlichen Ausgabe pro Probe eines Illumina-MiSeq-Laufes mit 384 parallel sequenzierten Proben; weitere „out-of-the-box“ Lösungen sind verfügbar: Mothur, QIIME.

<sup>2</sup> d = dies, Tag; h = hora, Stunde

<sup>3</sup> GB = Gigabyte

Weitere Erklärungen finden sich im Text und im zusammengefassten Abkürzungsverzeichnis am Ende des Buchs.

Auch die Datenerfassung gestaltet sich bisher sehr heterogen. Für die Nachvollziehbarkeit von wissenschaftlichen Ergebnissen müssen alle relevanten Metadaten wie beispielsweise Probenaufbereitungen, Verarbeitungsmerkmale, Sequenzierungsplattformen und Analysemethoden bekannt sein und standardisiert dokumentiert werden. Nur dann können Daten aus verschiedenen Quellen oder von verschiedenen Erhebungszeitpunkten in Relation gesetzt oder für weitere Analysen als gemeinsame Datenbasis verwendet werden [13, 14]. Anders als bei der Annotation von Biomaterial-bezogenen Informationen erfolgt die Erfassung von Omics-Daten bisher wenig standardisiert. Mit der *Minimum Information about a (Meta)Genome Sequence* (MIGS) [15] des *Genomic Standards Consortium* [15] und dem Schema *Sequence Read Archive* (SRA) des *European Nucleotide Archive* wurden zwar erste standardisierte Beschreibungen für DNA-Sequenzen entwickelt. In der Praxis finden jedoch beide Annotationsschemata wenig Anwendung. Weiterhin sind die *Minimum Information About a Microarray Experiment* (MIAME) für Microarray-Daten [16], die *Minimum Information About a Proteomics Experiment* (MIAPE) [17, 18], und die *Metabolomics Standards Initiative* (MSI) zu erwähnen. Die derzeit unzureichende Nutzung der Standards, sowohl in der akademischen Forschung als auch in der Pharmaindustrie, führt dazu, dass die Ergebnisse oftmals kaum nachvollziehbar und damit nicht reproduzierbar sind [19, 20], was hinsichtlich der Interpretierbarkeit der Ergebnisse ein ernstes Problem darstellt.

Während Labor-Informationen-Management-Systeme (LIMS) im Biobanken-Sektor in den letzten Jahren zunehmende Verbreitung gefunden haben<sup>5</sup>, sind in Omics-Laboren entsprechende Software-Systeme nur in Einzelfällen in den Workflow integriert. Begründet wird dies durch das Fehlen von adäquaten Lösungen, die nicht nur Laborworkflows bei der Generierung von molekularen Hochdurchsatz-Daten unterstützen, sondern auch das Zusammenspiel von Proben- und Datenmanagement sowie die Projektdokumentation und Verwaltung der Auftraggeber adressieren. Derzeit können vermehrt Bestrebungen für die Entwicklung und Einführung derartiger Systeme beobachtet werden.<sup>6</sup>

Mittelfristig ist damit zu rechnen, dass Dienstleistungen in den Omics-Disziplinen aus Kosten- und Qualitätsgründen weiter professionalisiert und zentralisiert werden. Dabei spielt die Gestaltung der Schnittstelle zwischen dem Auftraggeber und dem Dienstleister bei der Übermittlung der Ergebnisdaten eine sehr wichtige Rolle. Derzeit werden Genomics-Daten dem Auftraggeber vorwiegend im FASTQ-Format über (S)FTP- bzw. Webserver bereitgestellt. Auch nach der Übermittlung halten viele Dienstleister die Daten über einen längeren Zeitraum vor. Mit dem steigenden Einsatz von Omics-Technologien werden Dienstleister dazu jedoch kaum mehr in der Lage sein.

---

5 Siehe auch das Kapitel zu Biobanken im vorliegenden Bericht.

6 Die Beobachtungen basieren auf Experteninterviews und Vor-Ort-Analysen im Projekt GenoPerspektiv.

Vielmehr müssen Verfahren bestimmt und etabliert werden, wer die Daten nach den Regeln der guten wissenschaftlichen Praxis aufbewahren muss.

### **Ethische und rechtliche Rahmenbedingungen**

Um das Potential neuer Technologien für die patientennahe und translationale Forschung und Versorgung möglichst weitgehend zu nutzen, müssen die ethischen und rechtlichen Konsequenzen möglichst frühzeitig erkannt und adressiert werden. Dies wird beispielsweise in dem BMBF-geförderten Projekt „GenoPerspektiv“<sup>7</sup> aus der ELSA-Initiative und im Rahmen des Projekts „EURAT“<sup>8</sup> untersucht. Aufgrund der spezifischen Natur und besonderen Sensibilität genetischer Daten stellen sich in der Genomforschung ethische und rechtliche Fragen in besonderer Form. Die besondere Sensibilität genetischer Daten besteht darin, dass sich zahlreiche Informationen über Geschlecht, Herkunft und Krankheitsdispositionen aus dem Genom ableiten lassen. Das Informationspotenzial nimmt mit der fortlaufenden Erforschung des Genoms und seiner Rolle für Krankheiten und Persönlichkeitsmerkmale beständig zu. Darüber hinaus gelten genetische Daten aufgrund ihrer besonderen Persönlichkeitsrelevanz, ihrer „Schicksalshaftigkeit“ und der zwangsläufigen Mitbetroffenheit genetisch Verwandter als besonders sensibel. Um ihrer Sensibilität gerecht zu werden und beispielsweise einer möglichen sozialen Stigmatisierung oder Diskriminierung zu begegnen, unterliegen diese Daten einem besonderen Schutzbedarf.

Die ethischen und rechtlichen Herausforderungen, die sich durch die Sequenzierung des Genoms (WGS) oder von Teilen des Genoms (WES) ergeben, betreffen vor allem a) den Datenschutz und die Vereinbarkeit von Datenschutz einerseits und Forschungsbedürfnissen andererseits, b) die Gestaltung des Aufklärungs- und Einwilligungsverfahrens, c) die Handhabung von Zusatzbefunden und Forschungsergebnissen mit persönlicher Gesundheitsrelevanz für die Probanden und d) die Verantwortung nicht-ärztlicher Forscher. All diese ethischen und rechtlichen Herausforderungen sind eng miteinander verwoben und verweisen in vielen Punkten aufeinander [21]. Im Folgenden werden die aufgeführten Punkte näher beleuchtet:

- a) Um die Identität der Probanden zu verbergen und damit einen versehentlichen oder gar absichtlichen Missbrauch von genetischen Daten zu verhindern, werden Mechanismen der Pseudonymisierung oder Anonymisierung eingesetzt. Die Pseudonymisierung hat den Vorteil, dass es grundsätzlich möglich ist, Ergebnisse aus der Forschung, die für den Probanden gesundheitlich wichtig sein könnten (sogenannte

---

7 [www.genoperspektiv.de](http://www.genoperspektiv.de)

8 <http://www.uni-heidelberg.de/totalsequenzierung/>

Zufalls-, Zusatz- oder Forschungsbefunde), an den Probanden zurückzumelden. Eine absolute Anonymisierung kann es jedoch im Umgang mit individuellen genetischen Daten im Regelfall aufgrund des inhärenten Personenbezugs nicht geben. Sowohl der abnehmende Aufwand, eine Referenzprobe zu gewinnen und einen Abgleich der genetischen Daten durchzuführen, als auch die zunehmenden Möglichkeiten, nach in den genetischen Daten kodierten Eigenschaften im Internet oder in sozialen Netzwerken zu recherchieren, führen dazu, dass heute auch von einer faktischen Anonymisierbarkeit genetischer Daten kaum noch ausgegangen werden kann. Die Re-Identifizierung eines Probanden ist vor allem dann möglich, wenn ein Proband oder einer seiner Verwandten genetische Informationen zusammen mit dem Namen oder anderen Verbindungen zur Person öffentlich oder schlecht geschützt ins Internet gestellt hat [22]. Generell gibt es mehrere Wege, auf denen es (teilweise erwiesenermaßen) möglich ist, dank genetischer Daten Menschen zu re-identifizieren, die eigentlich anonym bleiben möchten [23]. Eine besondere Herausforderung für den Schutz der Vertraulichkeit der Daten und der Identität des Probanden wird ausgerechnet durch das erzeugt, was die Genomdaten für die biomedizinische Forschung erst wirklich interessant werden lässt: das Zusammenführen bzw. die Kombination der genomischen Daten mit phänotypischen bzw. biometrischen Daten, klinischen Verlaufsdaten oder Lebensstildaten.

Im Umgang mit Daten aus molekularen Hochdurchsatz-Analysen müssen zudem folgende Regeln und Richtlinien beachtet werden:

- ▶ Datenschutzgesetze des Bundes und der Länder;
- ▶ Landeskrankenhausgesetze (Sekundärnutzung von Daten aus der Krankenversorgung);
- ▶ Gute wissenschaftliche Praxis [13]. Für Aufbewahrungsfristen von 10 Jahren sollen die Daten auf haltbaren und gesicherten Trägern abgelegt werden. Als Grundlage von wissenschaftlichen Veröffentlichungen müssen die Daten vollständig und die wissenschaftlichen Ergebnisse nachvollziehbar dokumentiert sein.

Eine besondere Herausforderung für die Zukunft stellt die neue Datenschutzgrundverordnung dar, die im April 2016 verabschiedet wurde und ab Mai 2018 anwendbar sein wird. Bei der Planung von Projekten sollte im Voraus bedacht werden, wie diese neue EU-Datenschutzregelung zu befolgen ist. Offen ist allerdings noch, wie die bestehenden nationalen Gesetze an diese neue Verordnung angepasst werden und insbesondere, wie der nationale Gesetzgeber die in der Verordnung definierten Freiräume – gerade für den Umgang mit Gesundheitsdaten – für nationale Regelungen nutzen wird.

- b) Zentrale Voraussetzung der Erhebung und Nutzung molekularbiologischer Daten ist, wie in anderen Forschungsbereichen auch, die informierte Einwilligung (informed consent). Im Rahmen von Forschungsprojekten oder Studien müssen dazu dem Probanden alle notwendigen Informationen in verständlicher und klar formulierter Weise zur Verfügung gestellt werden. Um die Tragweite der Entscheidung überblicken zu können, muss die Aufklärung konkret und deutlich über Zweck, Umfang und Risiken bei der Datennutzung und -speicherung informieren. Aus ethischer Perspektive besteht ein entscheidendes Merkmal von genetischen Daten aus Hochdurchsatz-Analysen jedoch darin, dass im großen Umfang sensible Daten gesammelt, aufbereitet und interpretiert werden, deren Aussage, Tragweite und Verwendungsmöglichkeiten zum Zeitpunkt der informierten und selbstbestimmten Zustimmung nicht klar und schon gar nicht vollumfänglich abzusehen sind. Herausforderungen, die dabei auftreten, betreffen unter anderem den Umgang mit etwaigen Zusatzbefunden, die Form der Ergebnismeldung sowie den Umgang mit zu einem späteren Zeitpunkt auftretenden Interpretations- und Verwendungsmöglichkeiten der erhobenen Daten [24]. Eine systematische Einschätzung zum aktuellen Wissensstand, den moralischen Einstellungen sowie den Erwartungen und Ängsten der Probanden ist u. a. Gegenstand von diversen BMBF-Forschungsprojekten aus der aktuellen ELSA-Initiative des BMBF zur Systemmedizin.<sup>9</sup> Für die genomische Forschung mit Proben aus Biobanken liegen erste empirische Ergebnisse bezüglich der Einstellungen der involvierten Parteien zu einigen der genannten Herausforderungen vor. Sie besagen u. a., dass die Ergebnisse aus öffentlichen Umfragen („public“) und Experten-Interviews („professionals“) eher für eine Beschränkung der Rückmeldung auf klinisch behandelbare Ergebnisse sprechen, während die Patienten dazu neigen, alle Informationen mitgeteilt bekommen zu wollen [25].
- c) Damit kommen wir zur Problematik der individuell bedeutsamen Befunde aus der genomischen Forschung. Um die Privatsphäre des Einzelnen zu stärken und zu schützen, gilt das Recht auf informationelle Selbstbestimmung, das sich aus dem allgemeinen Persönlichkeitsrecht (Art. 2 Abs. 1 GG in Verbindung mit Art. 1 Abs. 1 GG) entwickelt hat. Obwohl das Recht auf informationelle Selbstbestimmung nicht explizit im Grundgesetz genannt wird, genießt dieses Recht Verfassungsrang. Es verleiht dem Individuum grundsätzlich die Befugnis, selbst darüber zu entscheiden, ob und in welchem Umfang persönliche Informationen offenbart werden.

Daraus ergibt sich sowohl das Recht auf Wissen als auch das Recht auf Nichtwissen [26]. Werden im Rahmen von Forschungsprojekten oder Studien genetische Daten

---

<sup>9</sup> Siehe <http://www.gesundheitsforschung-bmbf.de/de/5133.php>.

erhoben und analysiert, können grundsätzlich auch gesundheitsrelevante Befunde aufkommen, die innerhalb oder außerhalb der eigentlichen wissenschaftlichen Fragestellung stehen. Herausforderungen ergeben sich dabei, da Veranlagungen für Krankheiten angetroffen werden können, die beim Probanden noch nicht aufgetreten sind, erst seine Nachfahren treffen können oder für die derzeit keine Behandlungsmöglichkeiten existieren. Prinzipiell hat jeder Proband das Recht im Voraus im Laufe des Einwilligungsprozesses zu entscheiden, ob er über mögliche Befunde mit individueller Gesundheitsrelevanz informiert werden möchte oder nicht. Es gibt aber fast eine unbegrenzte Zahl möglicher Zufallsbefunde, da nicht nur unterschiedliche Krankheitsanlagen in Betracht kommen, sondern sich diese nochmals gemäß Penetranz und Expressivität unterscheiden. Um den Probanden ansatzweise aufzuklären und in die Lage zu versetzen, seine Präferenzen bezüglich der Rückmeldung gesundheitlich relevanter Informationen anzugeben, muss auf abstrakte Kategorien, in die sich mögliche Befunde ungefähr einteilen lassen, zurückgegriffen werden. Kriterien, um solche Kategorien von Befunden zu bilden, sind u. a. die Schwere der Krankheit, die Wahrscheinlichkeit ihres Auftretens (Penetranz), das Vorhandensein von Präventions- und Therapiemöglichkeiten und die Bedeutung für die Fortpflanzung [27, 28]. Sollte ein Proband darauf bestehen, über die gesamten genetischen Befunde der Analyse informiert zu werden, kann ihm dies aufgrund des Rechts auf informationelle Selbstbestimmung nicht versagt werden. Es ist allerdings noch weitgehend unklar, in welcher Form hier eine Aufklärung erfolgen kann, welche Qualitätsstandards eingehalten werden müssen und wie die breite Öffentlichkeit über den jeweils aktuellen Forschungsstand umfassend genug informiert werden kann. Es kann auch eine entsprechende Verzichtserklärung der Probanden zur Bedingung für die Teilnahme an einem Forschungsprojekt oder einer Studie gemacht werden. Diese ist jedoch immer widerrufbar [29].

- d) Die Problematik, dass Forscher im Laufe ihrer genomischen Forschungsstudie auf genetische Merkmale treffen können, deren Kenntnis für den Probanden gesundheitlich sehr bedeutsam sein kann und über die der Proband und seine Ärzte sehr wahrscheinlich keine Kenntnis haben, wirft die Frage nach den Verantwortlichkeiten nicht-ärztlicher Forscher auf [27]. Anders als die berufsrechtlich zum Wohl der Patienten verpflichteten ärztlichen Forscher sind nicht-ärztliche Forscher einerseits in erster Linie ihrer Forschung und der Gewinnung neuer wissenschaftlicher Erkenntnisse verpflichtet. Auf der anderen Seite rücken sie im Rahmen translationaler Forschung immer mehr an die klinische Seite heran und können z. B. bei ihrer Forschung auf genetische Merkmale treffen, deren Kenntnis für den Probanden wichtig ist, um sich beispielsweise präventiv gegen gefährliche Erbanlagen zu schützen. Die rechtliche Position des Forschers gegenüber Probanden

in genomischen Studien ist bezüglich dieser Problematik wohl weitgehend unbestimmt. Die Idee einer rechtlichen oder moralischen Pflicht des Forschers, aktiv nach möglichen bedeutsamen Merkmalen im Genom zu suchen, hat sich bisher nicht durchgesetzt. Weitgehend Übereinstimmung herrscht jedoch darüber, dass ein Forscher, der zufällig auf ein gesundheitlich sehr relevantes genetisches Merkmal stößt, dieses zurückmelden sollte [30]. Vor der Durchführung eines genomischen Forschungsprojekts sollte eruiert werden, wie wahrscheinlich das Auffinden von Merkmalen mit erheblicher Gesundheitsbedeutung für Probanden, innerhalb wie außerhalb der Forschungsfragestellung, sein wird, und wie genau derartige Merkmale validiert und als Befund zurückgemeldet werden sollten.

## Kategorisierung der IT-Landschaft

Die hier betrachteten Systeme zum Umgang mit Omics-Daten lassen sich generell in drei Kategorien einteilen:

1. IT-Systeme, die in einer zentralen Service-Einheit betrieben werden und Daten für Forscher und Ärzte exportieren (s. auch Review von Berger [31]);
2. Produkte an der Schnittstelle vom Datenproduzenten zum Auswerter (s. auch Review von Canuel [32]);
3. Lösungen für Omics-Daten mit Wissensmanagement-Perspektive.

In der ersten Kategorie werden aufgrund der hohen Bandbreitenanforderungen zwischen Messgerät und Primärspeicher oftmals lokale Server für Speicherung und Primärauswertung eingesetzt. Dort werden auch die entsprechenden Softwarelösungen betrieben.

Die Datenproduzenten werden meist nicht in der Lage sein, die Ergebnisdaten über einen längeren Zeitraum online vorhalten zu können. Vielmehr sind für die Übergabe der Daten etablierte Verfahren erforderlich, wie sie beispielsweise seit vielen Jahren in der klinischen Chemie (Ringversuche, QM-Zertifizierung) anzutreffen sind. In der Regel muss der Corresponding Author die Daten nach den Regeln guter wissenschaftlicher Praxis (good scientific practice, GSP) vorhalten.

In vielen Fällen gibt es keinen automatischen Austausch für Omics-Daten und deren Meta-Daten zwischen Klinik, Datenproduzenten, Medizinischer Informatik, Bioinformatik und Biostatistik. Klinische Meta-Daten werden oft nach wie vor in Excel-Tabellen abgelegt und transportiert. Daher ist die Schnittstelle vom Datenproduzenten zum Auswerter von hoher Relevanz. Großer Handlungsbedarf besteht hier, um die Reproduzierbarkeit und nachhaltige Verfügbarkeit der Daten gemäß GSP und guter klinischer

Praxis (*good clinical practice*; GCP) zu erreichen. In dieser Kategorie zeichnen sich erste Lösungen ab, die aus Initiativen wie FAIRDOM hervorgingen.

In der dritten Kategorie sind Systeme zu verzeichnen, mit denen die Daten weiter annotiert werden können. Anhand der Systeme können beispielsweise die konkreten Daten zur Formulierung von Abfragen an externe Informationsdienste aus der Systembiologie dienen und so in den Kontext eines spezifischen biologischen Pathway gestellt werden.

Für ein tiefgreifendes, systembiologisches Verständnis der komplexen Zusammenhänge im Bereich der Life Sciences und der klinischen Forschung ist die Verknüpfung mit weiteren IT-Infrastruktur-Komponenten essentiell. Dazu zählen vorrangig Biomaterialproben mit ihren Annotationsdaten, phänotypische Verlaufsdaten sowie weitere Probanden-bezogene Informationen. Auch weitere Ergebnisse aus nicht-molekularen Analysen oder Bilddaten können beteiligt sein. Für die Verknüpfung unterschiedlicher Datentypen werden Data-Warehouse-Lösungen eingesetzt.

Konzepte zur datenschutzgerechten Gestaltung von IT-Architekturen für die Integration von Omics-Daten in die verschiedenen Forschungslandschaften werden in [33] beschrieben.

## Aktueller Bestand

Zurzeit werden im Forschungskontext unterschiedliche organisatorische **Ansätze** verfolgt und verschiedene **Werkzeuge** eingesetzt. Trotz der derzeit sehr heterogenen Landschaft an den verschiedenen Standorten haben sich bestimmte Lösungen entlang der Omics-Pipelines etabliert. In der aktuellen Erhebung wird versucht, einen Überblick über die verfügbaren Software-Produkte zu geben. Aufgrund der z. T. schnelllebigen Open-Source-Softwarestacks erheben wir keinen Anspruch auf Vollständigkeit.

### 1. Produkte zum Einsatz beim Omics-Dienstleister (s. auch Review von Berger [31]):

#### 1.1. Analyse

- ▶ Bioconductor; Erweiterungen von R (<http://www.bioconductor.org>)
- ▶ BioDAS; Strukturbiologie (<http://www.biodas.org>)
- ▶ Bowtie; Alignmentwerkzeug (<http://bowtie-bio.sourceforge.net>)
- ▶ Burrows-Wheeler Aligner; Alignmentwerkzeug (<http://bio-bwa.sourceforge.net>)
- ▶ Chipster; Analyseplattform (<http://chipster.csc.fi>)
- ▶ CLC bio; Analyseplattform (<http://www.clcbio.com>)
- ▶ Cytoscape; Netzwerkanalyse (<http://www.cytoscape.org>)

- ▶ Galaxy; Pipeline Tool (<http://www.galaxyproject.org>)
- ▶ GenePattern; Analysewerkzeug (<http://www.broadinstitute.org/cancer/software/genepattern>)
- ▶ GeneXplain<sup>10</sup>; Portal mit Werkzeugen und Workflow-Management für Systembiologie und Cheminformatik; kann kommerzielle Datenbanken wie TRANSPATH und TRANSFAC mit einbinden.
- ▶ Genome Analysis Toolkit (GATK); Analysewerkzeug für Sequenzierung (<https://www.broadinstitute.org/gatk>)
- ▶ JMP Genomics (SAS) ([http://www.jmp.com/de\\_de/software/jmp-genomics.html](http://www.jmp.com/de_de/software/jmp-genomics.html))
- ▶ KNIME; Pipeline Tool (<http://www.knime.org/knime>)
- ▶ Picard; Werkzeug für Sequenzierungsdaten (<http://broadinstitute.github.io/picard>)
- ▶ R; Analyse (<http://www.r-project.org>)
- ▶ SAM Tools; Alignmentwerkzeuge (<http://samtools.sourceforge.net>)
- ▶ Short Oligonucleotide Analysis Package (SOAP); Alignmentwerkzeug (<http://soap.genomics.org.cn>)
- ▶ STRING; Datenbank für Proteininteraktion (<http://string-db.org>)
- ▶ Taverna; Analysewerkzeug (<http://www.taverna.org.uk>)

## 1.2. Ablage

- ▶ Addama; Datenbank für heterogene Daten (<http://www.systemsbio.org/addama>)
- ▶ ArrayExpress; Datenbank für funktionelle genomische Daten (<http://www.ebi.ac.uk/arrayexpress>)
- ▶ Biological General Repository for Interaction Datasets (BioGRID); (<http://thebiogrid.org/>)
- ▶ BioMart; Datenbanksystem für biologische Daten (<http://www.biomart.org>)
- ▶ BioModels Database; Datenbank für Modelle von biologischen Prozessen (<http://www.ebi.ac.uk/biomodels-main>)
- ▶ Epigenomesportal; Epigenetische Datenbank (<http://epigenomesportal.ca/ihec>)
- ▶ European Genome-phenome Archive (EGA); Datenbank für genomische Studien mit restriktivem Zugang (<https://www.ebi.ac.uk/ega>)
- ▶ European Nucleotide Archive (ENA); Datenbank für Nukleotidsequenzen (<https://www.ebi.ac.uk/ena>)
- ▶ IntAct; Molekulare Interaktions-Datenbank (<http://www.ebi.ac.uk/intact/>)
- ▶ iRODS; Integrated Rule-Oriented Data System; Open Source Data Management Software ([www.irods.org](http://www.irods.org))<sup>11</sup>
- ▶ openBIS [34]; offenes verteiltes System für das Management biologischer Daten ([www.cisid.ethz.ch/software/openBIS](http://www.cisid.ethz.ch/software/openBIS))

<sup>10</sup> <http://genexplain.com/genexplain-platform-1>

<sup>11</sup> Siehe auch Kapitel Forschungsdatenmanagement.

- ▶ PRIDE; Datenbank für Proteindaten (<http://www.ebi.ac.uk/pride>)
- ▶ Protein Data Base (PDB); Datenbank für Proteindaten (<http://www.wwpdb.org>)
- ▶ SDCubes; Datenbank für heterogene Daten (<http://www.semanticbiology.com/software/sdcube>)
- ▶ UniProt; Datenbank für Proteindaten (<http://www.uniprot.org>)

## **2. Produkte an der Schnittstelle vom Datenproduzenten zum Auswerter (s. auch Review von Canuel [32]):**

- ▶ 1000 Genomes; Datenpool für genomische Daten (<http://www.1000genomes.org/data>)
- ▶ ArrayExpress; (<http://www.ebi.ac.uk/arrayexpress/>)
- ▶ cBioPortal; Analysedatenbank für genetische Daten (<http://www.cbioportal.org>)
- ▶ Gene Expression Omnibus (GEO); (<http://www.ncbi.nlm.nih.gov/geo/>)
- ▶ ICGC Data Portal; Analysedatenbank für genetische und genomische Daten (<https://dcc.icgc.org>)
- ▶ openBIS [34]; offenes verteiltes System für das Management biologischer Daten ([www.cisd.ethz.ch/software/openBIS](http://www.cisd.ethz.ch/software/openBIS))
- ▶ SEEK [35]; Plattform zum Austausch von Daten und Modellen in der Systembiologie, soll von FAIRDOM mit openBIS in openSEEK zusammengeführt werden.
- ▶ The Cancer Genome Atlas (TCGA); (<http://cancergenome.nih.gov/>)
- ▶ The Cancer Proteome Atlas; (<http://bioinformatics.mdanderson.org/main/TCPA:Overview>)
- ▶ tranSMART [36]; ein zunächst aus i2b2 [37, 38] hervorgegangener Softwarestack der Transmart Foundation. tranSMART ist ein Werkzeug zur Verwaltung, Zusammenführung, Filterung und Abfrage von Daten aus der translationalen Forschung. Neben den phänotypischen Daten können mehrere Omics-Formate mit verwaltet werden. Derzeit prominenter Einsatz in Paris am HEGP [32].

## **3. Lösungen für Omics-Daten mit Wissensmanagement-Perspektive:**

- ▶ BioXM<sup>12</sup> [39, 40]; Umgebung zur Integration und Analyse biomedizinischer Daten in eine Wissensmanagement-Umgebung.
- ▶ eTRIKS<sup>13</sup> (European Translational Information and Knowledge Management Services); ergänzt tranSMART um weitere Services im Wissensmanagement. Die 16 Partner kommen aus der Wissenschaft und der Pharmaindustrie.
- ▶ OBO; Ontologien für Lebenswissenschaften (<http://www.obofoundry.org>)

---

12 <http://www.biomax.com/products/bioxm-knowledge-management-environment/>

13 <http://www.etriks.org/>

#### 4. Initiativen für besseres Datenmanagement in der biomedizinischen Informatik bzw. Systembiologie:

- ▶ de.NBI; das „deutsche Netzwerk für Bioinformatik-Infrastruktur“ bietet umfassende Dienstleistungen in den Lebenswissenschaften und der Biomedizin (<http://www.denbi.de>)
- ▶ ELIXIR; Europäische Infrastruktur für biologische Informationen zur Unterstützung der Lebenswissenschaften (<http://www.elixir-europe.org>)
- ▶ FAIRDOM<sup>14</sup>: „Support and Service Network for European Systems Biology“. Hierzu fand 2013 ein Workshop in Wien und 2014 im DKTK in Heidelberg statt. Das Produkt openBIS ging bereits aus der Initiative hervor. Ziel ist die Standardisierung, das Management und die Weitergabe von Daten und Datenmodellen in der Systembiologie (FAIR: Findable, Accessible, Interoperable and Reusable).
- ▶ Research Data Alliance (RDA)<sup>15</sup>; Ziel ist der schnellere Austausch von Forschungsdaten; bisher Schwerpunkt „Data Foundation and Terminology Working Group“.

#### Bereitstellung und Veröffentlichung von OMICS-Daten

In zentralen Forschungsrepositories sind bereits große Sammlungen an Genomen, epigenetischen Datensätzen, genetischen Polymorphismen oder Haplotypen, Gen-Expressionsdatensätzen basierend auf Microarrays oder RNA-Seq-Experimenten sowie auch von Proteomen und Proteinexpressionsdaten verfügbar. Forschungsdaten müssen in der Regel mit der Veröffentlichung in einem Journal mit Peer-Review-Verfahren in einer der großen Datenbanken hinterlegt werden. Im Fall von Genexpressionsdaten ist dies z. B. Gene Expression Omnibus (GEO), oder ArrayExpress, bei genetischen und epigenetischen Daten das „European Nucleotide Archive (ENA)“ für allgemeinen Zugriff oder das „European Genome-phenome Archive (EGA)“ für restriktiven Zugriff. Weiterhin gibt es eine Reihe spezialisierter Datenbanken, in denen die Daten aus größeren Sequenzierprojekten verfügbar sind, z. B. 1000 Genomes, The Cancer Genome Atlas (TCGA) oder das International Cancer Genome Consortium (ICGC). Für die Bereitstellung von Proteindatensätzen gibt es z. B. die „PROteomics IDentifications (PRIDE) database“ und „The Cancer Proteome Atlas“.

---

14 <http://www.fair-dom.org>

15 <http://europe.rd-alliance.org>

# Medizinisches Forschungsdatenmanagement an der CAU Kiel

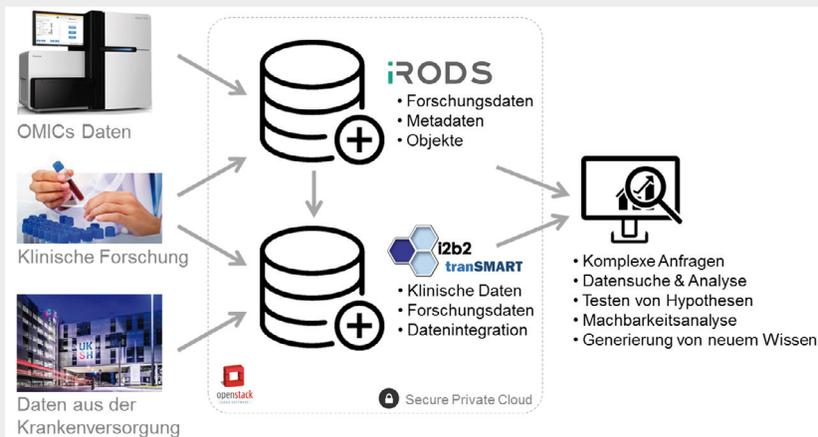


Abbildung 7.2: Medizinisches Forschungsdatenmanagement inkl. OMICS-Daten an der Christian-Albrechts-Universität zu Kiel (CAU).

Die Abbildung 7.2 zeigt das Konzept des klinischen Forschungsdatenmanagements an der Medizinischen Fakultät der Christian-Albrechts-Universität zu Kiel (CAU), welches derzeit in einem DFG-geförderten Referenzprojekt (§91b GG) zusammen mit Partnern aus der IT-Industrie und dem Universitätsklinikum Schleswig-Holstein (UK-SH) umgesetzt wird. Zielsetzung des Konzepts ist die Metadaten-gestützte Integration aller am Standort verfügbaren medizinischen Forschungsdaten, sowie den forschungsrelevanten klinischen Daten in pseudonymisierter oder anonymisierter Form aus dem Versorgungskontext. Darüber hinaus sollen mit Hilfe der geplanten Infrastrukturen den Wissenschaftlern generische Werkzeuge für nachhaltiges und DFG-konformes Datenmanagement (Speicherung und Archivierung) zur Verfügung gestellt werden.

Das Grundgerüst bildet eine „Secure Private Cloud“ Umgebung, die mittels OpenStack, einer Open-Source Virtualisierungsplattform implementiert wird. Über diese Lösung ist es möglich flexibel und nachhaltig auf die IT-Bedarfe der Nutzer (Rechenkapazität, Speicherplatz, Archiv) durch Anschluss verschiedener, skalierbarer IT-Ressourcen reagieren zu können. Hauptbestandteil der Installati-

on werden verschiedene sogenannte „Software-defined Storage“ Lösungen sein, die je nach Bedarf unterschiedliche Speicherarten (Block-, Objekt- oder Datei-Storage) bereitstellen. Schwerpunkt hierbei ist der neuartige Objekt-Speicher, in dem Daten nicht mehr in klassischen hierarchischen Dateisystemen auf dedizierten Datenträgern gehalten werden, sondern direkt über URLs (ähnlich den Internetseiten) referenzierbar und zugreifbar sind. Die für die Archivierung der Forschungsdaten benötigten Ressourcen werden durch eine vom Rechenzentrum der CAU betriebene Tape-Library zur Verfügung gestellt.

Für das Metadaten-gestützte Management der Forschungsdaten kommt als weitere Open-Source-Lösung iRODS (integrated rule-oriented data system) zum Einsatz, welche eine verteilte und automatisierbare Datenhaltung, sowie die Erfassung und Suche von flexibel definierbaren Metadaten ermöglicht. In dem entstehenden „Forschungsdaten-Repository“ werden neben studienbasierten Forschungsdaten auch alle Arten von OMICs-Daten aus den molekularbiologischen Laboren sowie insbesondere die klinischen Daten aus dem Versorgungskontext gesammelt, strukturiert und der Wissenschaft zur Verfügung gestellt. Um dem Nutzer die so gesammelten Daten möglichst einfach und übersichtlich zur Verfügung zu stellen oder zumindest über deren Existenz zu informieren, wird eine klinische Data-Warehouse Lösung (i2b2/transSMART) genutzt. Über ein Web-basiertes Frontend dieser Implementation können komplexe Abfragen sowie einfache, statistische Tests auf den verfügbaren Datenbeständen des Systems gemacht werden. So können Wissenschaftler zu individuellen Fragestellungen ermitteln, ob und wie viele Datensätze am Standort vorliegen und vom wem sie Zugriff auf diese Daten für ihre Forschungszwecke bekommen. Damit ermöglicht das System eine schnelle Machbarkeitsanalyse und das Testen von ersten Hypothesen.

Das ultimative Ziel ist die Generierung eines Mehrwerts für die Krankenversorgung. Denn neue Erkenntnisse, welche durch die Analyse von hoch komplexen Datensätzen (Stichwort: BigData), die über solch eine Datenintegrationsinfrastruktur zur Verfügung gestellt werden können, gewonnen werden, sollen letztlich auch in den klinischen Alltag einfließen (Translation) und so die Versorgung und Behandlung von Patienten verbessern.

## Bewertung und Handlungsbedarf

Omics-Methoden und -technologien haben das Potential, ein tiefgreifendes, systembiologisches Verständnis von komplexen Zusammenhängen im Bereich der Life Sciences und der klinischen Versorgung zu schaffen. Um den wissenschaftlichen Herausforderungen begegnen, nachhaltige Ressourcen aufbauen und Daten aus molekularen Hochdurchsatz-Analysen im Zusammenhang mit phänotypischen Informationen untersuchen zu können, kann folgender Handlungsbedarf identifiziert werden (siehe auch [41]):

- ▶ Geeignete Annotations-Schemata für die strukturierte und harmonisierte Erfassung von Daten aus Analysen des Epigenoms, Transkriptom, Proteoms und Metaboloms müssen entwickelt und etabliert werden.
- ▶ Bereits existierende Schemata wie MIGS und SRA für die Annotation von genomischen Daten müssen bezüglich ihres Einsatzes in der Praxis überprüft, bewertet und ggf. überarbeitet werden.
- ▶ Während professionelle Lösungen im Biobanking-Bereich für die Dokumentation der Laborworkflows und Projekte sowie für das Proben- und Datenmanagement bereits weit verbreitet sind, besteht weiterhin großer Bedarf an der Entwicklung und Einführung von adäquaten, standardisierten IT-Lösungen für Omics-Labore.
- ▶ Der Austausch von Omics-Daten und zugehöriger Meta-Daten zwischen Klinik, Datenproduzent, Biomedizinischer Informatik und Biostatistik erfolgt oft suboptimal und in den wenigsten Fällen automatisiert. Hier besteht sehr großer Handlungsbedarf, um die nachhaltige Verfügbarkeit der Daten gemäß GSP und GCP und eine Reproduzierbarkeit von Ergebnissen zu erreichen. Notwendig ist sowohl die Entwicklung und Etablierung entsprechender Standards als auch die Verfügbarkeit und Nutzung standardisierter IT-Unterstützung.
- ▶ Bei der Archivierung von Daten und Projektergebnissen sind neben den Rohdaten auch Informationen aus der Qualitätsanalyse, Arbeitsdaten aus den Laborprozessen sowie Angaben zum Projektdesign inkl. Probenlisten und Abrechnungsinformationen abzulegen und vorzuhalten.
- ▶ Rohdaten aus NGS-Analysen werden derzeit in der Regel im FASTQ-, bcl- oder BAM-Format aufbewahrt. Diese Formate bieten die Möglichkeit, dass alle weiteren Formate aus ihnen generiert werden können. Dazu müssen aber zusätzlich auch die entsprechenden Analysepipelines, Software-Systeme und die jeweils verwendeten Versionen der Genomdatenbanken aufgehoben werden. Daher wird meistens zusätzlich noch das Endergebnis der Analyse aufgehoben. Dies kann z. B. die Tabelle mit Transkriptanzahlen bei RNA-SEQ-Experimenten sein. Alternativ kommen auch VCF-Dateien mit analysierten SNP-Varianten bei genomischer Sequenzierung in Frage.

- ▶ Es ist festzustellen, dass bei weiter sinkenden Kosten für molekulare Hochdurchsatz-Analysen massive Probleme bei der langfristigen Speicherung der Daten auftreten werden. Um Doppelvorhaltungen bei Dienstleistern und Auftraggebern zu vermeiden, sind Verfahren und Regelungen zu etablieren, die klären, wer die Daten nach den Regeln der guten wissenschaftlichen Praxis aufbewahren muss.<sup>16</sup>
- ▶ Auswertung und Interpretation der Daten verlangen fundierte bioinformatische und biostatistische Kenntnisse sowie das Verständnis systembiologischer Zusammenhänge. Oftmals ist das Wissen von Forschern und Ärzten jedoch mangelhaft. Um Missstände zu beseitigen und künftigen Problemen vorzubeugen, muss der Umgang mit und die Analyse von molekularen Hochdurchsatz-Daten in der Ausbildung von Forschern und Ärzten verankert werden. Zudem sollten entsprechende Experten bei der Analyse solcher Daten in jedem Fall beteiligt sein.
- ▶ Allgemeine Übersichten über vorhandene Geräte und transparente (zentral gesteuerte) Vorgehensweisen bei der Anschaffung von Neu-Geräten an einem Standort sind erforderlich, um effiziente Auslastungen zu erzielen und unnötige Doppelvorhaltungen zu vermeiden.
- ▶ Der Aufbau und Erhalt nachhaltiger Infrastrukturen kann nicht auf der Basis projektbezogener Mittel erfolgen, sondern erfordert die Bereitschaft von wissenschaftlichen Forschungseinrichtungen, Institutionen des Gesundheitssystems sowie dem Bund und den Ländern zur Bereitstellung der notwendigen Mittel.
- ▶ Aus Kosten- und Qualitätsgründen ist mittelfristig damit zu rechnen, dass Dienstleistungen in den Omics-Disziplinen weiter professionalisiert und zentralisiert werden.
- ▶ Im Zuge der erforderlichen Standardisierung des Daten- und Prozessmanagements in der Systembiologie ist eine enge Abstimmung zwischen den verschiedenen Omics-Disziplinen und der Bioinformatik notwendig.
- ▶ Für den stetigen Betrieb von zentralen Service-Einrichtungen ist die Sicherung und Weiterentwicklung von Know-how und Hands-On-Expertise essentiell. Gefährdungen durch befristete Beschäftigungsverhältnisse im akademischen Mittelbau müssen abgefedert werden. Um genügend Expertise zu erhalten bzw. diese auszubauen ist zudem eine enge Anbindung an wissenschaftlich verwandte Einheiten nötig.
- ▶ Bisherige Infrastrukturen sind forschungsorientiert aufgebaut. Aufgrund der hohen Anforderungen in der Krankenversorgung in Bezug auf Qualität, Verfügbarkeit und Validität molekularer Hochdurchsatz-Daten sind dort eigene Strukturen erforderlich. Generische und breit einsetzbare Konzepte für die Integration von NGS-Analysen in die klinische Routine müssen entwickelt und umgesetzt werden. Diese müssen allerdings so konkret beschrieben sein, dass sie die Umsetzung auch praktisch unterstützen können.

---

<sup>16</sup> siehe Empfehlung 7 in [42, S. 21f]

- Ethische und rechtliche Fragestellungen, insbesondere im Umgang mit Zufallsbefunden und Befunden, die direkte Familienangehörige betreffen oder auf neuen wissenschaftlichen Erkenntnissen beruhen, müssen geklärt werden.

## Danksagung

Diese Arbeit wurde neben der Förderung durch die TMF<sup>17</sup> auch durch das BMBF im Rahmen des Forschungs- und Förderkonzeptes e:Med (Projekt „GenoPerspektiv“ aus der ELSA-Initiative, Förderkennzeichen 01GP1402 und Projekt „sysINFLAME“, Förderkennzeichen 01ZX1306C) unterstützt.

## Literatur

1. Ng, S.B., Nickerson, D.A., Bamshad, M.J., Shendure, J., *Massively parallel sequencing and rare disease*. Hum Mol Genet, 2010. **19**(R2): S. R119-124.
2. Pao, W., Hutchinson, K.E., *Chipping away at the lung cancer genome*. Nat Med, 2012. **18**(3): S. 349-351.
3. Biesecker, L.G., *The new world of clinical genomics*. J Clin Endocrinol Metab, 2012. **97**(11): S. 3912-3914.
4. Biesecker, L.G., *Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project*. Genet Med, 2012. **14**(4): S. 393-398.
5. Biesecker, L.G., Burke, W., Kohane, I., Plon, S.E., Zimmern, R., *Next-generation sequencing in the clinic: are we ready?* Nat Rev Genet, 2012. **13**(11): S. 818-824.
6. Bhatti, J., Redelmeier, D.A., *Angelina Jolie and medical decision science*. Med Decis Making, 2015. **35**(1): S. 4-5.
7. Evans, D.G., Barwell, J., Eccles, D.M. et al., *The Angelina Jolie effect: how high celebrity profile can have a major impact on provision of cancer related services*. Breast Cancer Res, 2014. **16**(5): S. 442.
8. James, P.A., Mitchell, G., Bogwitz, M., Lindeman, G.J., *The Angelina Jolie effect*. Med J Aust, 2013. **199**(10): S. 646.
9. Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L., Rice, P.M., *The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants*. Nucleic Acids Res, 2010. **38**(6): S. 1767-1771.

---

<sup>17</sup> Die Erstellung dieses IT-Reports durch die TMF wurde wiederum vom BMBF im Rahmen des Projekts MethInfraNet (FKZ 01GI1003) gefördert, s. auch einleitendes Kapitel.

10. Herrmann, A., Hampe, J. *Workshop Genomdaten - Ergebnisprotokoll (LABIMI/F)*. 2012, [http://www.labimi-f.med.uni-goettingen.de/Deliverables/D4\\_2\\_Workshop.pdf](http://www.labimi-f.med.uni-goettingen.de/Deliverables/D4_2_Workshop.pdf) (Abruf: 2016-11-21).
11. Li, H., Handsaker, B., Wysoker, A. et al., *The Sequence Alignment/Map format and SAMtools*. *Bioinformatics*, 2009. **25**(16): S. 2078-2079.
12. Danecek, P., Auton, A., Abecasis, G. et al., *The variant call format and VCFtools*. *Bioinformatics*, 2011. **27**(15): S. 2156-2158.
13. DFG *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*. 2009. Deutsche Forschungsgemeinschaft, Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, Unterausschuss für Informationsmanagement, [http://dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf) (Abruf: 2016-11-21).
14. Dickmann, F., Grütz, R., Rienhoff, O., *A "Meta"-Perspective on "Bit Rot" of Biomedical Research Data - Quality of Life through Quality of Information*. *Studies in Health Technology and Informatics*, 2013. **180**: S. 260 - 264.
15. Field, D., Garrity, G., Gray, T. et al., *The minimum information about a genome sequence (MIGS) specification*. *Nat Biotechnol*, 2008. **26**(5): S. 541-547.
16. Brazma, A., Hingamp, P., Quackenbush, J. et al., *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data*. *Nat Genet*, 2001. **29**(4): S. 365-371.
17. Taylor, C.F., *Minimum reporting requirements for proteomics: a MIAPE primer*. *Proteomics*, 2006. **6 Suppl 2**: S. 39-44.
18. Taylor, C.F., Paton, N.W., Lilley, K.S. et al., *The minimum information about a proteomics experiment (MIAPE)*. *Nat Biotechnol*, 2007. **25**(8): S. 887-893.
19. McDermott, J.E., *Reproducibility: two more red flags for suspect work*. *Nature*, 2013. **499**(7458): S. 284.
20. Begley, C.G., *Six red flags for suspect work*. *Nature*, 2013. **497**(7450): S. 433 - 434.
21. Winkler, E.C., Schickhardt, C., *Ethische Herausforderungen der Genomsequenzierung in der translationalen Forschung und Antworten aus dem EURAT-Projekt*. *LaboratoriumsMedizin*, 2014. **38**(4): S. 211-220.
22. Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y., *Identifying Personal Genomes by Surname Inference*. *Science*, 2013. **339**(6117): S. 321 - 324.
23. Erlich, Y., Narayanan, A., *Routes for breaching and protecting genetic privacy*. *Nat Rev Genet*, 2014. **15**(6): S. 409-421.
24. 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.

25. Husedzinovic, A., Ose, D., Schickhardt, C., Fröhling, S., Winkler, E.C., *Stakeholders' perspectives on biobank-based genomic research: systematic review of the literature*. Eur J Hum Genet, 2015. **23**(12): S. 1607-1614.
26. Duttge, G., *Das Recht auf Nichtwissen in der Medizin*. Datenschutz und Datensicherheit, 2010. **2010**(1): S. 34 - 38.
27. EURAT *Eckpunkte für eine Heidelberger Praxis der Ganzgenomsequenzierung* 2013. Projektgruppe Ethische und Rechtliche Aspekte der Totalsequenzierung des menschlichen Genoms (EURAT) [http://www.marsilius-kolleg.uni-heidelberg.de/md/einrichtungen/mk/presse/stellungnahme\\_heidelberger\\_praxis\\_der\\_ganzgenomsequenzierungl\\_2013-06-12.pdf](http://www.marsilius-kolleg.uni-heidelberg.de/md/einrichtungen/mk/presse/stellungnahme_heidelberger_praxis_der_ganzgenomsequenzierungl_2013-06-12.pdf) (Abruf: 2016-11-21).
28. Rudnik-Schöneborn, S., Langanke, M., Erdmann, P., Robiński, J., *Ethische und rechtliche Aspekte im Umgang mit genetischen Zufallsbefunden – Herausforderungen und Lösungsansätze*. Ethik in der Medizin, 2013. **26**(2): S. 105-119.
29. NER *Biobanken für die Forschung*. 2004. Nationaler Ethikrat, <http://www.ethikrat.org/dateien/pdf/biobanken-fuer-die-forschung.pdf> (Abruf: 2016-11-21).
30. Bredenoord, A.L., Kroes, H.Y., Cuppen, E., Parker, M., van Delden, J.J., *Disclosure of individual genetic data to research participants: the debate reconsidered*. Trends Genet, 2011. **27**(2): S. 41-47.
31. Berger, B., Peng, J., Singh, M., *Computational solutions for omics data*. Nat Rev Genet, 2013. **14**(5): S. 333-346.
32. Canuel, V., Rance, B., Avillach, P., Degoulet, P., Burgun, A., *Translational research platforms integrating clinical and omics data: a review of publicly available solutions*. Brief Bioinform, 2014.
33. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
34. Bauch, A., Adamczyk, I., Buczek, P. et al., *openBIS: a flexible framework for managing and analyzing complex data in biology research*. BMC Bioinformatics, 2011. **12**: S. 468.
35. Wolstencroft, K., Owen, S., du Preez, F., Krebs, O., Mueller, W., Goble, C., Snoep, J.L., *The SEEK: a platform for sharing data and models in systems biology*. Methods Enzymol, 2011. **500**: S. 629-655.
36. Athey, B.D., Braxenthaler, M., Haas, M., Guo, Y., *transSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research*. AMIA Jt Summits Transl Sci Proc, 2013. **2013**: S. 6-8.
37. Segagni, D., Tibollo, V., Dagliati, A., Zambelli, A., Priori, S.G., Bellazzi, R., *An ICT infrastructure to integrate clinical and molecular data in oncology research*. BMC Bioinformatics, 2012. **13 Suppl 4**: S. S5.

38. Kohane, I.S., Churchill, S.E., Murphy, S.N., *A translational engine at the national scale: informatics for integrating biology and the bedside*. J Am Med Inform Assoc, 2012. **19**(2): S. 181-185.
39. Cano, I., Tenyi, A., Schueller, C. et al., *The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research*. J Transl Med, 2014. **12 Suppl 2**: S. S6.
40. Maier, D., Kalus, W., Wolff, M. et al., *Knowledge management for systems biology a general and visually driven framework applied to translational medicine*. BMC Syst Biol, 2011. **5**: S. 38.
41. Umbach, N., Löhnhardt, B., Sax, U., *Managing OMICS-Data: Considerations for the Design of a Clinical Research IT-Infrastructure*. Stud Health Technol Inform, 2015. **216**: S. 668-671.
42. DFG, *Denkschrift: Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“*. 2 Aufl. 2013, Wiley-VCH, Weinheim, [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf) (Abruf: 2016-11-21).



# Mobile IT-Werkzeuge

*Karoline Buckow, Henning Dathe, Thorsten Rottmann, Otto Rienhoff*

## Anwendungsbereich und Eingrenzung

### Mobile IT-Werkzeuge im Gesundheitswesen

Ein IT-Werkzeug, das „medizinische Studien revolutionieren [...] und vielleicht sogar die Medizin für immer verändern“<sup>1</sup> kann – so beschrieb Apple 2015 sein ‚ResearchKit‘, das Millionen iPhones und andere Smartphones weltweit in Erfassungsinstrumente für medizinische Forschungsprojekte verwandeln wollte. Ein Jahr später stellte Apple seinem Open Source Framework das ‚CareKit‘ zur Seite, welches Bürger und Patienten individuell in die eigenen Versorgungsprozesse sowie Disease Management einbinden soll. In der Selbstdarstellung des Konzerns ist die im Vorjahr angekündigte Revolution bereits fest im wissenschaftlichen und klinischen Alltag angekommen. Entgegen dieser ambitionierten Selbsteinschätzung finden sich aktuell im großen amerikanischen Studienregister<sup>2</sup> zwar 230 klinische Studien zum Thema ‚Apple‘ in Amerika, aber nur drei zum Stichwort ‚ResearchKit‘ und eine Studie zum Begriff ‚Research Kit‘. Der Durchbruch dieser Technologie steht also noch bevor – zumindest in der von Studienregistern erfassten und eher regulierten klinischen Forschung. Die Idee der mobilen Datenerfassung oder Datenaufzeichnung ist zudem nicht neu: IT-Werkzeuge zum Messen und Übertragen medizinischer Daten sind seit Jahrzehnten Bestandteil des Gesundheitswesens. Erste telemedizinische Ansätze basierten in der Regel auf einer telefonischen Datenübermittlung [1]. Für das Kompetenznetz Demenzen wurde bereits vor zehn Jahren die Möglichkeit der mobilen Dateneingabe über UMTS geprüft. Dieser Versuch musste aufgrund der zu geringen Netzabdeckung eingestellt werden.<sup>3</sup> Der seitdem erfolgte Ausbau der Netzwerkinfrastruktur, die Entwicklung besserer Techniken und Standards für die Mobilkommunikation, Fortschritte in den Bereichen Mikrotechnologie, Datenverarbeitung, Datenspeicherung und Akkuleistung sowie die Entwicklung innovativer mobiler Mess- und Telekommunikationsgeräte erlauben es, zunehmend mobile IT-Werkzeuge in der medizinischen Versorgung und Forschung einzusetzen, etwa zu Verlaufsbeobachtungen [2]. Dabei entstehen zum Teil grundsätzlich neue Methoden, mit denen Prozesse quantifiziert werden können. Als Beispiel sei hier die chemische und physikalische Vermessung chronischer Wunden genannt

---

1 <http://www.apple.com/de/researchkit/>, Stand 2015

2 <https://clinicaltrials.gov/>, Stand: 21.9.2016

3 Eine genauere Darstellung dieses und anderer „früher“ Projekte bot ein TMF-Workshop aus dem Jahr 2005, siehe [www.tmf-ev.de/news/368](http://www.tmf-ev.de/news/368).

oder das breite Spektrum von Lab-on-a-Chip-Anwendungen, die sich zunehmend in der ambulanten Versorgung und der medizinischen Forschung zu etablieren beginnen [3]. Dazu können auch epidemiologisch orientierte Messsysteme gezählt werden, die z. B. das Auftreten von Mückenlarven in Feuchtgebieten aufzeichnen oder etwa Infektionssymptome in Menschengruppen automatisch messen. Mobile IT-Lösungen für die beschriebenen Anwendungsszenarien reichen von Sensoren, die ohne Interaktion Daten des Untersuchungsobjektes übermitteln (z. B. implantierte oder textile Sensoren, die Vitalparameter erfassen und übertragen), bis hin zu Smartphones und Smartwatches, die mit dem Anwender unter Berücksichtigung aktueller Lokalisations- und Bewegungsinformationen interagieren können. Neue Aufmerksamkeit gewinnt auch eine neue Generation von Closed-loop-Systemen – vor allem in der Intensivmedizin.

Aktuelle Smartphone- und Tablet-PC-Modelle erfahren eine breite Akzeptanz, unter anderem aufgrund ihrer auch für Laien intuitiv nutzbaren Bedienoberfläche und der breiten Verfügbarkeit von günstigen Tarifen zur mobilen Datenübertragung. Sowohl die inzwischen weite Verbreitung dieser sogenannten mobilen Devices als auch die verfügbaren Schnittstellen erlauben einen vielseitigen Einsatz. Aufgrund dieser beachtlichen Entwicklung und ihres erheblichen Potenzials konzentriert sich dieser Bericht exemplarisch auf Applikationen (Apps), die im Kontext der Gesundheitsforschung für Smartphones oder Tablet-PCs verfügbar sind oder entwickelt werden. Dabei werden auch mobile Devices berücksichtigt, die mittels Sensoren Daten aufzeichnen und an eine entsprechende Applikation, onboard auf dem Device selbst (beispielsweise HealthKit) oder offboard in eine Cloud-Infrastruktur (beispielsweise Google Fit), übertragen.

### **Einsatzbereiche von Apps im Gesundheitswesen**

Mit der rasant steigenden Anzahl und Verbreitung von Smartphones oder Tablets steigt auch die Anzahl von Anwendungen, die die neuen Funktionen und Techniken des Mobile Computing nutzen und anwendbar machen. Aktuell werden mehr als 165.000 Apps im Apple App Store und Google Play Store von den Herstellern der Kategorie mHealth zugeordnet [4]. Sie bedienen nahezu das gesamte Spektrum von Dienstleistungen im Gesundheitswesen. Hierzu zählen:

- ▶ die Promotion von Anbietern des Gesundheitswesens, z. B. Empfehlungen zur Einhaltung bekannter Diätpläne, Trainingsanleitungen bekannter Fitness-Anbieter, Produktfinder oder Anwendungshilfen für bestimmte Marken oder Produkte;
- ▶ die Prävention von Krankheiten, z. B. Übungsanleitungen für das Training der Rückenmuskulatur, Aufzeichnung gelaufener Strecken mittels GPS, Erstellung von Ernährungsplänen und die Beobachtung von deren Einhaltung;

- ▶ die Bereitstellung medizinischer Informationen, z. B. Klinikleitfäden, medizinische Wörterbücher, Kataloge für Arzneimittelinformationen;
- ▶ das Selbstmanagement von Nutzern des Gesundheitssystems, z. B. Patiententagebücher, Erinnerung und Dokumentation für Medikamenteneinnahmen, Erinnerung an fällige Arztbesuche und Impfungen;
- ▶ das Krankheits-Monitoring, z. B. mobile Systeme für ein personalisiertes Health Care Monitoring, mobile Aufzeichnung von Vitalparametern oder Umweltinformationen über Sensoren;
- ▶ die Unterstützung von medizinischem Fachpersonal bei der Versorgung von Patienten, z. B. Protokollierung von Notfalleinsätzen [5], Betrachtung von Bilddaten, Kommunikation und Informationsübermittlung zwischen verschiedenen Akteuren der Patientenversorgung;
- ▶ die Bereitstellung von Informationen zur Rehabilitation, z. B. Suchmaschinen für Reha-Einrichtungen, Informationen und Checklisten für Reha-Aufenthalte.

Unabhängig vom jeweiligen Einsatzbereich, dient ein Großteil der entwickelten Applikationen dazu, persönliche Gesundheitsdaten der Anwender zu sammeln – und die Bereitschaft, private Informationen über die Gesundheit, Fitness oder Lifestyle zu teilen, scheint sehr ausgeprägt zu sein [2]. Für das enorme Ausmaß des hier entstandenen und weiter wachsenden Marktes ist die rasante Zunahme von Apps im Bereich mHealth über die letzten Jahre ein deutlicher Indikator.<sup>4</sup>

### **Einsatz von Apps für die patientenorientierte Datenerfassung**

Zunehmend lassen sich Apps auch in der Forschung einsetzen, zum Beispiel um im Bereich der personalisierten Medizin eine am Probanden orientierte Datenerfassung zu unterstützen. Mit dem jüngst vorgestellten „CareKit“ möchte Apple Patienten die Möglichkeit geben, ihre persönlichen Geräte bidirektional in die Infrastruktur ihres Versorgers zu integrieren, und so zu einem Akteur in der Versorgungs-IT zu werden. Die Erfassung von Informationen, die sich nicht über regelmäßige (z. B. halbjährliche) Konsultationen im Rahmen von Studien-Visitenplänen dokumentieren lassen, stellt neue Anforderungen an den methodischen Umgang mit den gewonnenen Daten. Die Berücksichtigung tagesabhängiger Umweltfaktoren, individueller Lebensqualitätsdaten oder im Tagesverlauf schwankender Symptomatik erfordert die Partizipation des Patienten. Dabei sollen Patienten die Lücken von Studien-Visitenplänen durch die Erfassung eigener Beobachtungen oder individueller Outcome-Parameter füllen und zu einer Vervollständigung des Gesamtbildes über einen definierten Beobachtungszeitraum

<sup>4</sup> Dossiers zu Markt- und weiteren Aspekten findet man bei <http://de.statista.com/> unter den Stichworten „Digital Health“ oder präziser zum Thema „mHealth“.

hinweg beitragen. Die unmittelbare Aufzeichnung von Erlebtem und von Wahrnehmungen soll einen Bias durch Gedächtniseffekte vermeiden, der typischerweise darin besteht, dass die Erinnerungen zu positiv oder in Abhängigkeit vom Ausgang des Ereignisses oder des aktuellen Befindens dargestellt werden [6, 7].

Mit einer solchen Erfassung sogenannter Patient Reported Outcomes (PRO) sollen Probleme von Patienten bezüglich des Befindens oder des Zurechtkommens im Alltag früh erkannt, Schlussfolgerungen auf erforderliche therapeutische Maßnahmen ermöglicht und eine Verbesserung der Lebensqualität und Zufriedenheit mit der Gesundheitsversorgung erreicht werden [8]. Als PRO werden laut der Arzneimittelzulassungsbehörde der USA, der Food and Drug Administration (FDA), direkt vom Patienten bezogene Statusinformationen über dessen Gesundheitszustand definiert – eine Interpretation dieses Gesundheitsberichts durch den Arzt oder anderweitiges medizinisches Fachpersonal ist dabei explizit ausgeschlossen [9]. Mobile Geräte wie Smartphones oder Tablets und die entsprechenden Applikationen können aufgrund ihrer zunehmenden Verbreitung ein geeignetes Werkzeug für die patientenorientierte Datenerfassung sein. In diesem Bericht werden daher mobile IT-Werkzeuge fokussiert, die dem Patienten die eigenständige Erfassung von PRO erlauben und den Patienten somit aktiv in die Datenerfassung für Forschungsvorhaben einbinden.

Einsatzgebiete mobiler Sensoren sind vielfältig und umfassen unter anderem die Aufnahme von Aktivitätsdaten, Schlafmustern, Blutwerten und Vitalparametern. Die Sensoren sind dabei in Abhängigkeit von ihrem Einsatzgebiet entweder in ein Smartphone bzw. ein spezielles separates Device integriert oder sie werden am oder im Körper getragen [10]. Und der Markt für Devices und Dienste, die mobile Sensoren einsetzen und verarbeiten, steigt – mit erheblichen Auswirkungen auf die damit verbundene Datenmenge [10, 11]. Gleichzeitig sind jedoch auch die Erwartungen groß, in bislang nicht erreichbare Informationswelten vorstoßen zu können, die die Umwelt oder den Lebensstil eines Probanden beschreiben. Diese Daten könnten Zusammenhänge zwischen Lebensgewohnheiten und dem Ausbrechen einer Krankheit aufzeigen [11].

Das Potential der Nutzung persönlicher Gesundheitsdaten für die medizinische Forschung – erfasst über mobile Devices – hat auch Apple erkannt und veröffentlichte im März 2015 das ResearchKit, ein Open Source Framework zur Entwicklung von Apps für die Sammlung großer Datenmengen für medizinische Forschungsprojekte [12]. Über Apps, deren Entwicklung auf dem ResearchKit basiert, können Daten von dem Patienten und über den Patienten erfasst und an eine zentrale Datenhaltungsstelle für Auswertungen transferiert werden: Daten können dabei aus Fragebögen und anhand spezieller On-board-Sensoren (Mikrofon, Bildschirm z. B. für Tipp-Übungen, Be-

schleunigungsmesser und Gyroskop z. B. zur Erfassung von Bewegungs- und Lageinformationen) aufgezeichnet werden [13, 14]. Ziel ist es, größere und heterogenere Studienpopulationen als bisher in Forschungsprojekte einzuschließen und anhand der Einbeziehung von Umwelt-, Aktivitäts- und Interaktionsdaten Aufschlüsse über den Einfluss bestimmter Parameter und Rahmenbedingungen auf die Gesundheit zu gewinnen [15].

## Betreiber, Anwender, Nutzer

Zielgruppe für die Verwertung der Daten, die aus der patientenorientierten Datenerfassung im Forschungskontext hervorgehen, sind in der Regel Forschungseinrichtungen aus dem akademischen Bereich, der pharmazeutischen Industrie oder Institutionen des Gesundheitswesens, wie Leistungserbringer oder Kostenträger. Zunehmend wird auch der Anwender selbst zum Nutzer seiner selbsterfassten Daten, da mobile Applikationen häufig über Visualisierungsfunktionen verfügen und dem Anwender so Ergebnisse oder Statistiken direkt verfügbar machen. Äquivalent zu bestehender Standard-Software für die Datenerfassung in klinischen Studien können die Konzeption, die (Weiter-)Entwicklung sowie die Bereitstellung der Applikation und die Speicherung der erfassten Daten als kommerzieller Service oder im Rahmen einer forschungsbezogenen Kollaboration angeboten werden. Betrachtet man die Gesamtheit der aktuell verfügbaren Apps, die dem Thema Health-Apps untergeordnet sind, werden über 60 Prozent von Einzelpersonen oder Kleinunternehmen mit unter zehn Mitarbeitern bereitgestellt [11]. Diese Situation führt zu Konsequenzen hinsichtlich der Nachhaltigkeit, Professionalität und Interoperabilität bestehender Entwicklungen. Plattformen wie Google Fit<sup>5</sup> oder Frameworks wie HealthKit<sup>6</sup> von Apple sammeln persönliche Gesundheitsdaten an einem Ort und bieten Softwareentwicklern entsprechende Programmierschnittstellen an, um Daten zwischen den Anwendungen auszutauschen, sofern die entsprechenden Zugriffsberechtigungen – entweder durch den Anwender oder mit Hilfe unveränderbarer Voreinstellungen – eingeräumt wurden.

## Rahmenbedingungen

### Regulatorische Rahmenbedingungen

Die Durchführung medizinischer Forschungsvorhaben einschließlich der Prozesse für die elektronische Erfassung und Verarbeitung von Daten im Rahmen medizinischer Forschung unterliegt den Vorgaben der Guten Klinischen Praxis (GCP), der Guten

---

5 <https://developers.google.com/fit/>

6 <https://developer.apple.com/healthkit/>

Wissenschaftlichen Praxis (GSP) sowie der Einhaltung nationaler wie internationaler Regeln und Gesetze. Glaubwürdige und qualitativ hochwertige Forschungsergebnisse erfordern eine Datenerfassung und -verarbeitung entsprechend geltender regulatorischer Rahmenbedingungen. Relevante Anforderungen werden hierbei durch die ICH GCP Guideline [16], den FDA Code of Federal Regulations 21 CFR Part 11 [17] sowie die FDA Guidance for Industry - Computerized Systems Used in Clinical Investigations [18] definiert. Die für die Konzeption und die Entwicklung von Apps erforderlichen Anforderungen umfassen Maßnahmen und Techniken zur Sicherstellung von Datenschutz und Datensicherheit, zur Unterstützung der Datenerfassung, zur Etablierung von Datensicherungsmechanismen, zur Datenkonsistenz und Datenintegrität sowie zur Gewährleistung der Wiedergewinnung von Daten [19]. Ebenso müssen Vorgaben zum Datenschutz eingehalten werden, um die Daten des App-Anwenders vor unbefugtem Zugriff zu schützen.

### **Technische Rahmenbedingungen**

Technische Anforderungen an den Einsatz von Apps auf mobilen Geräten für Forschungszwecke müssen auf verschiedenen Ebenen berücksichtigt werden:

- ▶ Anforderungen an die Benutzeroberfläche: Die Benutzeroberfläche muss relevante Funktions- und Designanforderungen erfüllen, um eine intuitive Bedienung zu gewährleisten. Dabei müssen gegebenenfalls spezifische Studienstrukturen wie Visitenpläne oder bestimmte Erfassungsmethoden dargestellt werden können. App-Entwickler müssen dabei ggf. unterschiedlich große Displays und Auflösungen beachten. Abhängig vom Einsatz der Anwendung sind Aspekte der Barrierefreiheit zu berücksichtigen.
- ▶ Anforderungen an die Kommunikation und Interoperabilität: Für eine Kommunikation der App mit weiteren Forschungsinfrastrukturkomponenten müssen Software-seitig Schnittstellenanforderungen erfüllt werden. Je nach Fragestellung des Forschungsprojekts sind ggf. weitere Schnittstellen erforderlich, die z. B. die Kommunikation mit Messgeräten oder Sensoren unterstützen. Beim Einsatz mobiler IT-Werkzeuge im ländlichen Raum ist zudem eine eventuell nicht vollständige Netzabdeckung zu bedenken.

- ▶ Anforderungen an die Datensicherheit sowie den Datenschutz: Datensicherheits- und Datenschutzerfordernungen beziehen sich sowohl auf die Sicherheit der auf dem Gerät des Studienteilnehmers gespeicherten Daten als auch auf die Sicherheit der Datenübertragung, zum Beispiel durch entsprechende Verschlüsselungstechniken oder Methoden der Nutzer-Authentifizierung. Dabei ist zu beachten, dass eine App z. B. auf dem privaten Smartphone eines Versuchsteilnehmers in einer unsicheren Umgebung läuft und dass eine persönliche Authentifizierung eines Versuchsteilnehmers nicht die durchgängige Pseudonymisierung seiner Daten aufhebt.
- ▶ Akzeptanzanforderungen: Die Akzeptanz einer App wird durch den Nutzer der App bestimmt, der bewusst oder unbewusst Anforderungen an das Gerät und die Anwendung stellt. Ein wichtiger Aspekt im Hinblick auf die Akzeptanz ist die Usability.
- ▶ Anforderungen an die Kosten für Entwicklung, Wartung und Betrieb: Sowohl für die Entwicklung als auch für die Wartung müssen unter anderem Aspekte der Plattformabhängigkeit sowie die erwartete Häufigkeit der Aktualisierung der Inhalte berücksichtigt werden.
- ▶ Anforderungen an die Metadatendokumentation: Die neuen Möglichkeiten der patientenorientierten Datenerfassung anhand einer unüberschaubaren Bandbreite an Devices bergen neue Gefahren für Fehlerhebungen oder geräte- bzw. prozessabhängige Fehler. Anders als im kontrollierten Umfeld von Studienzentren, lassen sich die Bedingungen der Datenerfassung oder Messung durch den Patienten am eigenen Smartphone deutlich schlechter standardisieren und eventuelle Umgebungseinflüsse sind schwer zu kontrollieren [20]. Deshalb müssen die Verfahren diesen Qualitätsaspekt nachvollziehbar adressieren.

### **Anwendungsfall- / Einsatz-abhängige Rahmenbedingungen**

Sowohl das im Rahmen eines Forschungsprojektes betrachtete Szenario als auch das damit verbundene Kollektiv implizieren bestimmte Bedingungen, die für den Einsatz der App in dem jeweiligen Umfeld betrachtet werden müssen. Bleiben limitierende Faktoren bezüglich des Einsatzes mobiler Technologien für die Datenerfassung unbeachtet, kann dies eine unerwünschte beziehungsweise ungeplante Selektion der Anwender zur Folge haben. Relevant hierfür können eventuelle Barrieren im Umgang mit der mobilen Datenerfassung sein (z. B. kein eigenes Smartphone, kein geübter Umgang mit dem Smartphone, kein mobiler Internetzugriff, Sorgen oder Ängste vor der Preisgabe persönlicher Informationen). Ebenso können krankheitsspezifische Besonderheiten (z. B. Tremor, Sehschwäche) einzelne Probanden oder das gesamte Kollektiv in der Anwendung der Studien-App beeinträchtigen. Entsprechende Einflüsse unterschiedlicher Kulturen, Probandengruppen (z. B. Altersgruppen) und auch zeitliche Trends (z. B. Hersteller-spezifische Open- und Closed-Source Frameworks) müssen bereits

im Studiendesign und zusätzlich im Rahmen der Auswertung (z. B. Adjustierung) berücksichtigt werden [13, 20].

Je nach Einsatz der App und in Abhängigkeit vom gewählten Einsatzbereich müssen unter Umständen außer dem Verbraucherschutz das Arzneimittelgesetz oder das Medizinproduktegesetz bei der Konzeption der App berücksichtigt werden. Die Europäische Kommission hat in einer im Januar 2012 veröffentlichten und im Juli 2016 aktualisierten Guideline Entscheidungskriterien definiert, unter welchen Bedingungen „stand alone software“ einschließlich Applikationen für mobile Geräte als Medizinprodukte (engl. medical devices) zu klassifizieren sind [21]. Auf nationaler Ebene hat das Bundesinstitut für Arzneimittel und Medizinprodukte in 2015 eine Orientierungshilfe erstellt.<sup>7</sup>

## Kategorisierung der IT-Landschaft

Es sind verschiedene Ebenen zur Klassifizierung der IT-Landschaft denkbar. So werden in der CHARISMHA-Studie neun Typen von Akteuren<sup>8</sup> und ebenfalls neun mögliche Handlungsfelder<sup>9</sup> unterschieden [22]. Diese werden u. a. tabellarisch gegeneinander aufgetragen, um dann die Handlungsoptionen innerhalb der entstandenen Zellen zu diskutieren. Zur Klassifizierung von Sensoren werden eigenständige, komplexe Hierarchien vorgeschlagen [23, 24]. Und ein umfassendes Klassifikationsverfahren für Digital-Health-Anwendungen wurde auch in einer von der Bertelsmann-Stiftung in Auftrag gegebenen Studie [25] entwickelt.

Bei der Einteilung der IT-Landschaft werden hier im Folgenden jedoch nur zwei verschiedene Ebenen der Betrachtung unterschieden: Die Darstellung der unter die mobilen IT-Werkzeuge fallenden Devices und Applikationen und die Darstellung der Forschungs-IT-Infrastruktur, in die Apps oder Devices integriert werden.

Die Gruppe der mobilen IT-Werkzeuge lässt sich in drei Kategorien unterteilen. Die erste Kategorie bilden mobile Devices wie Handys, Smartphones oder Tablets, die über entsprechende Apps Gesundheitsdienste oder -funktionen anbieten. Die zweite Kategorie bilden PDAs (persönliche digitale Assistenten) – tragbare Rechner, die zunehmend an Bedeutung gegenüber mobilen Devices verlieren. Die dritte Kategorie umfasst mobile IT-Werkzeuge, die am oder im Körper getragen werden. Hierzu zählen unter anderem Smartwatches, Wearables oder Implantate, die Messwerte über Sensoren aufzeichnen und versenden.

---

7 [http://www.bfarm.de/DE/Medizinprodukte/Abgrenzung/medical\\_apps/\\_node.html](http://www.bfarm.de/DE/Medizinprodukte/Abgrenzung/medical_apps/_node.html)

8 Bund/ Länder, Interessenvertretungen, Hersteller, Anbieter, Laien-Anwender, Profi-Anwender, Forscher, Versorgungseinrichtungen und Kostenträger.

9 Organisation und Infrastruktur, Finanzierung, Zugang, Ethik, Regulation, Forschung, Qualität, Transparenz und Information.

Für eine Verwertung von mobil erfassten Daten für Forschungszwecke ist eine Integration der datenaufnehmenden Applikation in bzw. eine Interoperabilität mit einer bestehenden validierten Studieninfrastruktur erforderlich. Diese Anbindung umfasst sowohl die Abbildung relevanter Studienstrukturen, studienrelevanter Standards und administrativer Studieninformationen in der App als auch die Verknüpfung der von den Patienten erfassten Daten mit den klinischen Daten der Studiensoftware über ein gemeinsames Identitätsmanagement unter Einhaltung geltender Datenschutzregelungen.

Ausgehend von einer Applikation zur Erfassung von PRO-Daten über mobile Devices, müssen Strukturen für die Datenerfassung vorgegeben werden, um diese durch die erfassten Daten patientenübergreifend auswertbar zu machen. Entsprechend müssen Maßnahmen etabliert werden, die die Einhaltung des Studiendesigns bestmöglich sicherstellen und Abweichungen dokumentieren. So ist unter Umständen die Vorgabe fixer Beobachtungszeitpunkte relevant für die Aussagekraft der erfassten Daten. Je nach Forschungsfragestellung müssen entsprechende Strukturen in der Studien-App abgebildet und der Patient muss an Erfassungstermine sowie studienspezifische und administrative Besonderheiten erinnert werden können. Zusätzlich zu vorgegebenen Erfassungszeitpunkten kann eine ereignisbasierte Dokumentation unterstützt werden. Die erforderlichen Projektstrukturen sowie der Rahmen, in dem Konfigurationen durch den Patienten in der App durchgeführt werden können (z. B. das Verschieben von Erfassungsterminen oder das Einstellen von Erinnerungs- oder Messzeiten), müssen in der App hinterlegt werden.

Die Auswertung der vom Patienten erfassten Informationen erfordert die Übertragung der PRO-Daten an eine zentrale Studiensoftware, die entsprechende Schnittstellen für die statistische Auswertung bereitstellt. Die Verknüpfung der in der Studien-App erfassten Daten mit eventuell verfügbaren vom Arzt dokumentierten Daten in der Studiendatenbank setzt ein übergreifendes Identitätsmanagement voraus. Entsprechende Schnittstellen und Prozesse, die für die Kommunikation zwischen Studiensoftware und einem projektspezifischen Identitätsmanagement bereits in vielen Forschungsverbänden etabliert sind, müssen für die Anbindung von Studien-Apps implementiert und getestet werden. Dabei muss die Übertragung der vom Patienten dokumentierten Daten in die Studiendatenbank konform mit geltenden Datenschutzregelungen erfolgen.

Der Einsatzbereich der Studien-App legt fest, ob und welche Geräteschnittstellen notwendig sind. Ist im Rahmen des Forschungsprojekts neben der Datenerfassung durch den Patienten eine Aufnahme von Daten aus externen Messinstrumenten oder von Sensoren erforderlich, müssen entsprechende Geräteschnittstellen implementiert, getestet und validiert werden.

## Aktueller Stand

### Relevanz von PRO-Daten für die klinische Forschung

Bereits im Jahr 2009 veröffentlichte die FDA eine Guidance for Industry zum Thema Patient Reported Outcome (PRO). Darin fordert sie eine stärkere Verwendung von PRO, um

- ▶ unterschiedliche Wahrnehmungen gesundheitsrelevanter Aspekte von Arzt und Patienten erkennen zu können,
- ▶ durch den Erinnerungs-Bias verschwommene Schilderungen zu vermeiden und
- ▶ die Reliabilität von Informationen durch Standardisierung (im Vergleich zu Arzt-geführten Interviews) zu optimieren [6].

Inhalte der Guidance sind unter anderem Empfehlungen für die Auswahl, Konzeption und Validierung von PRO-Instrumenten, für den Einsatz von Erfassungsmethoden, zum Einsatz von PRO in klinischen Studien sowie zum Umgang mit und zur Auswertung von PRO-Daten [9]. Dabei geht die FDA gesondert auf den Gebrauch elektronischer Erfassungsmethoden ein und weist auf die Gültigkeit der Forderungen des FDA Codes of Federal Regulations 21 Part 11 [17] auch für die elektronische Erfassung von PRO-Daten über mobile Geräte hin [9].

Im April 2014 veröffentlichte die Europäische Kommission ein Grünbuch zum Thema mobile Health-Dienste [11]. Das Dokument stellt das besondere Potential von mHealth für die Gesundheitsforschung und die Gesundheitsversorgung heraus, zeigt jedoch bestehende Hindernisse und Probleme beim Ausbau der mHealth-Dienste auf. Die hierzu aufgeworfenen Fragen machen einen Klärungs- bzw. Handlungsbedarf auf unterschiedlichen Ebenen deutlich. Dies betrifft neben anderen Themen die Einhaltung geltender Datenschutzvorschriften, die Transparenz über die Verarbeitung und Speicherung der Gesundheitsdaten einschließlich Datensicherheit, Zugriffsregelungen und Leistungsnachweisen, die Maßnahmen zur Prüfung von mHealth-Diensten und die Optimierung der Interoperabilität von mHealth-Diensten [11].

Im Juni 2015 fand eine Pressekonferenz zum Release einer von der Techniker Krankenkasse in Auftrag gegebenen Studie statt, die vom Studienzentrum der Universität Freiburg durchgeführt wurde [26]. Diese Studie fokussiert zunächst auf die Notwendigkeit der Entwicklung von Qualitätsstandards. Nach einer Markübersicht inklusive einer Checkliste für Verbraucher werden dann Aspekte von Technik, Datenschutz und Informationssicherheit kritisch diskutiert. Ein Fazit der Studie ist die Notwendigkeit

einer wissenschaftlichen Evaluation hinsichtlich des Nutzens von Gesundheits-Apps, wobei die Methoden zur Nutzenbewertung noch zu entwickeln sind.

Im Januar 2016 äußerte sich auch der Sachverständigenrat für Verbraucherfragen zu dem Thema [27]. Die Veröffentlichung hat eine vergleichbare Zielsetzung wie die bereits erwähnte CHARISMHA-Studie vom April 2016 [22]. Beide betonen neben den Chancen auch die möglichen Risiken von eHealth und mHealth. Den Vorteilen bei der Erfassung von Daten mit Apps stellt die CHARISMHA-Studie eine mangelnde Standardisierung der Apps sowie der zugehörigen Studiendesigns und Klärungsbedarf bei ethischen Fragen gegenüber.

### **Erfassung von PRO-Daten**

Aktuelle Publikationen zeigen die Vorteile einer PRO-Datenerfassung mit Hilfe elektronischer Medien gegenüber einer papierbasierten Datenerfassung auf. Hierzu zählen die Reduzierung fehlender Angaben [28], die Verhinderung invalider oder unvollständiger Eingaben durch die Möglichkeit der gesteuerten Datenerfassung (z. B. über Sprungmarken) [9, 28], die Eingrenzung der Antwortoptionen (z. B. über Radio-Buttons, durch die nur eine Antwortoption zugelassen wird) zur Vermeidung mehrdeutiger Antworten [28], die Steigerung der Compliance durch die Protokollierung sämtlicher Aktionen in einem Audit Trail [28, 29] sowie die Optimierung des Datenmanagements und der Datenqualität [30]. Tablet-PCs vereinigen die Vorteile papierbasierter Dokumentation in Bezug auf die Einfachheit einerseits und elektronischer Datenerfassung in Bezug auf das Datenmanagement und die Datenqualität andererseits, so das Ergebnis einer vergleichenden Studie über verschiedene Methoden der PRO-Datenerfassung [31]. Sie punkten vor allem in Bezug auf die Einfachheit der Anwendung, die Flexibilität und die Genauigkeit bzw. geringere Fehlerraten gegenüber anderen Methoden [31]. Mit der fortschreitenden Entwicklung entsprechender Erfassungsmethoden ist in den vergangenen Jahren ein Trend von papierbasierter Dokumentation (einfache oder mit Texterkennung mittels Optical Character Recognition) über telefonische Erfassungsmethoden (Interview oder mit Interactive Voice Response) hin zu einer Erfassung über PDAs, Tablet-PCs oder Smartphones gerade für die Erfassung außerhalb des klinischen Einsatzes zu erkennen [32].

Die zunehmende Verbreitung von Apps für die Erfassung persönlicher Gesundheitsdaten hat eine Entwicklung angestoßen, immer mehr Menschen zur Aufzeichnung ihrer Gesundheitsdaten zu bewegen. Dies gilt vor allem für Menschen mit chronischen Erkrankungen, die über einen langen Zeitraum die Entwicklung ihrer Erkrankung und entsprechende Begleiterscheinungen beobachten wollen. So entsteht eine persönliche

Datensammlung, die – alle diese persönlichen Datensammlungen zusammengenommen – einen erheblichen wissenschaftlichen Wert haben kann [13]. Die inzwischen weite Verbreitung von Smartphones und Tablet-PCs fördert diese Entwicklung und bietet neue Möglichkeiten der Partizipation des Patienten und der Berücksichtigung relevanter Einflüsse, auch im Bereich der Forschung. Die Datenerfassung durch den Studienteilnehmer kann über eine App auf entsprechenden mobilen Endgeräten zeitlich und räumlich flexibel gestaltet werden. Durch die patientenorientierte Dokumentation der Lebensumstände, der Lebensqualität oder der Lebensgewohnheiten entstehen Optionen, Therapieerfolge oder potentielle Einflussfaktoren auf eine Erkrankung erfassen zu können. Dabei können zusätzliche Parameter wie Lokalisationsdaten, Bewegungs- oder Aktivitätsprofile sowie Informationen über das Verhalten einer Person im Netz in die Analyse eingeschlossen werden.

Personalisierte Medizin und Patient Empowerment sind aktuelle „Bewegungen“ in der medizinischen Forschung, die einer geeigneten IT-Infrastruktur bedürfen. Der Einsatz mobiler Technologien kann hierfür eine wichtige technische Grundlage liefern – die Eignung von Apps für dieses Anwendungsfeld muss daher geprüft werden.

### **Vernetzung zum Thema Mobile Computing in Deutschland**

In Deutschland bietet die Arbeitsgruppe MoCoMed (Mobiles Computing in der Medizin) der Gesellschaft für Informatik e.V. (GI) und der Deutschen Gesellschaft für medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS) eine Plattform für den fachlichen Austausch und Kooperationsmöglichkeiten ihrer Mitglieder. Inhaltliche Schwerpunkte der Arbeitsgruppe sind dabei unter anderem innovative mobile Anwendungen im Gesundheitswesen, die Evaluation dieser Informationssysteme sowie telemedizinischer und telepflegerischer Dienstleistungen und die Betrachtung relevanter Grundlagen.<sup>10</sup> International gibt es die Wearable Sensors in Healthcare Working Group der IMIA zum Thema Sensorik.<sup>11</sup>

### **Datenschutzaspekte und Sicherheit von mHealth-Diensten**

Über Apps auf Smartphones oder Tablet-PCs kann eine patientenbasierte Datenerfassung von Lebensumständen und Umwelteinflüssen ermöglicht werden. Neben den Informationen, die der Patient selbst dokumentiert, können unter anderem durch eine GPS-Ortung der aktuelle Aufenthaltsort, Wetterbedingungen, Aktivitätsdaten und Bewegungsmuster erfasst, gespeichert und übermittelt werden [33]. Gerade bei den

---

<sup>10</sup> <http://fg-mocomed.gi.de/>

<sup>11</sup> <http://www.wearable-sensors.org/website/>

im Hintergrund laufenden Datenaufzeichnungen besteht die Gefahr einer fehlenden zeitlichen Begrenzung der Meldung von Informationen, bei gleichzeitig fehlendem Bewusstsein der Probanden in Bezug auf die fortdauernde Aufzeichnung [13]. Zudem kann eine Re-Identifizierung bei steigender Anzahl involvierter Messparameter immer weniger ausgeschlossen werden, verknüpft man beispielsweise eine spezifische Erkrankung mit Bewegungsprofilen [14, 34]. Die Akzeptanz der Probanden für den Einsatz mobiler IT-Systeme ist in der Regel von dem Verständnis für die Studieninhalte, der Erkenntlichkeit eines praktischen Nutzens sowie auch von der Integration des Systems in den Alltag abhängig [35]. Aber auch die Datensicherheit und der Datenschutz spielen eine zunehmende Rolle. Patienten müssen über die Inhalte der Studie sowie über die Art und den Umfang der Datengewinnung aufgeklärt werden [16]. Da die Datenerfassung im Alltag des Patienten als Eindringen in den privaten Raum empfunden werden könnte, sind eine genaue Aufklärung über mögliche Situationen während der Studie und die Durchführung eines Interviews nach Beendigung der Datenerhebung von Bedeutung [35]. Darüber hinaus muss sichergestellt werden, dass Minderjährige nicht ohne das Einverständnis ihrer Eltern an entsprechenden Datenerfassungen teilnehmen – eine Schwierigkeit, die durch die Verwendung des ResearchKit von Apple softwareseitig nicht gelöst wird [34]. Das Framework ist darauf ausgerichtet, die Einwilligung direkt über die App abzubilden. Ein Umgehen möglicher Sicherheitsfragen nach dem Alter ist damit relativ einfach möglich. Genauso ist es nur beschränkt möglich, sicherzustellen, dass die App-Benutzer wirklich für die Teilnahme an der Studie geeignet sind und dass Patienten über sämtliche Risiken der Studie und den Umgang mit ihren Daten informiert werden – ein persönliches Aufklärungsgespräch ist durch die App kaum ersetzbar [13, 20, 34].

Grundsätzlich gilt, dass die Nutzung von Daten aus mHealth-Diensten für die Forschung konform mit geltenden Regularien erfolgen muss. Für Forschungsvorhaben auf Basis personenbezogener Daten, die ursprünglich zu einem anderen Zweck erfasst wurden, ist eine ggf. fehlende Einwilligung durch den Patienten problematisch [11]. Transparenz bezüglich Datenverwertung, Datenversand, Datennutzung oder Datenbereitstellung ist erforderlich, um das Vertrauen der App-Anwender zu gewinnen. Ansätze wie „App synopsis“ oder „Apps Peer Review“ sind hierfür ein Schritt in die richtige Richtung [33]. Darüber hinaus mangelt es bislang an standardisierten Lösungen für den Nachweis über die Funktionstüchtigkeit entsprechender Apps, die Korrektheit der Funktionsweise bzw. über die Durchführung einer angemessenen Testung [11].

Im Rahmen eines TMF-Workshops 2010 wurde eine erste Rahmenarchitektur für AAL- und telemedizinische Anwendungen konzipiert, die hinsichtlich der notwendigen technischen und organisatorischen Maßnahmen zum Datenschutz drei Bereiche

unterscheidet: einen Heimbereich im direkten Lebensumfeld der betroffenen Patienten, einen Dienstleistungs- und Behandlungsbereich sowie einen Bereich für die ggf. sekundäre Nutzung der Daten zu Forschungszwecken. Der Dienstleistungsbereich vermittelt dabei zwischen den beiden anderen Bereichen [36]. Während der Übergang vom Dienstleistungs- zum Sekundärnutzungsbereich weitgehend gemäß der technischen und organisatorischen Vorgaben der generischen Datenschutzkonzepte der TMF [37] gestaltet werden kann, gibt es für den Heimbereich und dessen Übergang zum Dienstleistungsbereich bisher noch keine abgestimmten und breit akzeptierten Konzepte und Standards. Klar ist aber, dass für den Übergang ein Gateway mit einer Kontrollmöglichkeit für die Betroffenen benötigt wird [36].

In einem Folgeworkshop Anfang 2015 wurde deutlich, dass mit der Nutzung von Smartphones im Heimbereich und der damit im Regelfall verbundenen Integration der genutzten Apps und ihrer Datenverwaltung in eines der kommerziell ausgerichteten Ökosysteme der Firmen Apple, Google oder Microsoft häufig auch ein datenschutzrechtlich problematischer Kontrollverlust hinsichtlich des Datenflusses einher geht.<sup>12</sup>

Seit Juni 2016 gibt es für den Schutz von Gesundheitsdaten auf europäischer Ebene freiwillige praktische Verhaltensregeln für mHealth-Entwickler [38]. Diese umfassen neben den Hauptprinzipien Zweckbegrenzung, Datensparsamkeit, Datenschutz und Rechte an den Daten noch zehn konkrete weitere Punkte. Es wird sich zeigen, inwieweit diese Regeln bei den verschiedenen Herstellern umgesetzt werden.

Im September 2016 erschien eine im Auftrag der Techniker Krankenkasse durchgeführte Studie des IGES-Instituts [39]. Im Kern enthält sie den Vorschlag einer risikobasierten vierstufigen Einteilung von Apps hinsichtlich der Anforderungen zu ihrer Zulassung. Diese Risikoklassen sind, geordnet nach steigendem Risiko: Information, Sammlung von Daten, Unterstützung des Nutzers und Ersetzen der Leistungserbringer. In den zwei höchsten Risikoklassen sind Nachweise zu Wirksamkeit und Sicherheit zu erbringen. Fehlen solche Nachweise, wird ein Verbot des Vertriebs auf dem deutschen Markt gefordert ([39], Kap. 4.3).

### **Statistisch methodische Probleme**

Der Anspruch, jedem Patienten eine individuell optimale Therapie zukommen zu lassen, stellt nicht nur die Medizin vor neue Herausforderungen – auch die Methodik zur Erhebung, Verarbeitung und Auswertung von Daten steht wachsenden Anforderungen

---

<sup>12</sup> [www.tmf-ev.de/news/1690](http://www.tmf-ev.de/news/1690)

gegenüber, um Zusammenhänge individueller Faktoren und klinischer Outcome-Parameter im Rahmen personalisierter Forschungsvorhaben abbilden zu können.

Der Umgang mit PRO-Daten im Forschungskontext unterliegt grundsätzlich denselben Regeln wie sie auch für ärztlich erfasste Daten gelten [9]. Die FDA-Guidance für den Umgang mit PRO-Daten definiert, dass dies Überlegungen zur statistischen Auswertung von PRO-Daten einschließt [9]. Dennoch weist die FDA auf besondere Herausforderungen bei der statistischen Auswertung hin. Ein besonderer Aspekt hierbei sei der Umgang mit Missing Values. Der Umgang mit fehlenden Einzelangaben sollte dabei idealerweise schon bei der Entwicklung eines PRO-Instruments berücksichtigt und im Statistical Analysis Plan für die Studie spezifiziert werden [9]. Apps bieten hier die Möglichkeit, den Probanden zu definierten Zeitpunkten an die Datenerfassung zu erinnern oder auf fehlende Angaben hinzuweisen, um die Zahl fehlender Werte gering zu halten.

Die Reliabilität und die Validität eines Messverfahrens haben großen Einfluss auf die Qualität der Studienergebnisse. Üblicherweise werden Verfahren wie die Korrelation wiederholter Testergebnisse in die Reliabilitätsbestimmung und damit auch in die Auswahl der in einen Test einzuschließenden Items einbezogen. Dieses Verfahren lässt sich jedoch nicht problemlos für die wiederholte Datenerfassung am Patienten einsetzen, da hier gerade die Veränderungen zwischen unterschiedlichen Messzeitpunkten von Interesse sind [40]. Aufgrund der Bandbreite an Sensoren in mobilen Devices für die Aufzeichnung persönlicher Gesundheitsdaten sind die Validität und die Reliabilität der resultierenden Daten zu hinterfragen – dies wird umso deutlicher, betrachtet man den aufwändigen Zulassungsprozess für Medizinprodukte im Vergleich zu dem Veröffentlichungsprozess einer App [2].

Schwierig zu beantworten bleiben die Fragen nach der Qualität der Daten und damit bei Sensoren solche nach der Genauigkeit. Jede mit einem Sensor verbundene App verfügt über eine oder mehrere verdeckte Zwischenebene der Signalverarbeitung, welche dem Endnutzer oder Forscher nicht zugänglich sind. Erst in der App präsentiert sich das Signal als konkret verwertbares Endprodukt. Solange die Information über die Funktionsweise der intermediären Signalverarbeitung typischerweise ein Betriebsgeheimnis ist, kann auch keine punktgenaue Kalibration des Sensors erfolgen. Ein beispielhafte Darstellung dieser Problematik anhand eines Wearable findet sich in [41]. Im Notfall müsste für die Forschung der Weg eines Re-Engineerings in Form eines Nachbaus der Signalverarbeitung beschritten werden.

Ferner stellt diese intermediäre Signalverarbeitung die Verfügbarkeit von Rohdaten in Frage. Diese sind, etwa in Form von Visualisierungen oder Kennzahlen, den profes-

sionellen Anwendern aus dem Umgang mit Medizinprodukten durchaus vertraut. Hier besteht noch Klärungsbedarf bezüglich der notwendigerweise von Apps und Herstellern zur Verfügung zu stellenden Datenarten. Eine Prüfung des Einsatzes von Apps muss für jede Studie separat erfolgen. Akzeptanz, Verfügbarkeit sowie die psychologische und soziale Einstellung zu der Erfassungsmethode können von Studienkollektiv zu Studienkollektiv schwanken. Eine Nicht-Berücksichtigung dieser Faktoren könnte einen Selektions-Bias zur Folge haben und somit die Validität des Studienergebnisses gefährden.

## Bewertung und Handlungsbedarf

Apps und Sensoren sind vielversprechende Werkzeuge für die patientenorientierte Datenerfassung. Sie vereinen die Vorteile der computergestützten Datenerfassung mit einer intuitiven Nutzeroberfläche und nur von der Akkukapazität begrenzter Verfügbarkeit. Ein professioneller Einsatz im Rahmen wissenschaftlicher Forschung erfordert jedoch eine eingehende Evaluation des Ansatzes. Für die Interpretation der erfassten Daten ist ein Verständnis des Umgangs des Patienten mit dieser Technologie im Vergleich zu anderen Erfassungsmethoden notwendig. Zudem kann eben die Art dieses Umgangs von Studieninhalten und der Zielgruppe abhängig sein. Aus diesem Grund sollten für den Einsatz von Apps zur PRO-Datenerfassung die technologie- und zielgruppenspezifische Eignung der Methodik sowie anwendungsfallbezogene Anforderungen geprüft werden.

Der Einsatz von Apps im GCP-konformen Studienumfeld erfordert eine technisch validierte Anbindung an eine zertifizierte Studiensoftware. Dies bedeutet die Abbildung relevanter Studienstrukturen und -standards sowie administrativer Studieninformationen in der App. Die vom Patienten erfassten Daten müssen mit den klinischen Daten der Studiensoftware über ein gemeinsames Identitätsmanagement unter Einhaltung geltender Datenschutzregelungen und Rahmenbedingungen verknüpft werden. Weiterhin müssen die Reliabilität, die Validität und die Relevanz der PRO-Daten bewertet werden, um den Nutzen und die Auswirkungen auf die medizinische Forschung beurteilen zu können. Aufgrund der intransparenten Dritteinflüsse durch Hersteller von Betriebssoftware (maßgeblich iOS und Android) sowie der Firmware der vielzähligen Sensorgeräte gestalten sich das Nachvollziehen der Datenintegrität sowie die Feststellung der Datenqualität derzeit schwierig.

Eine zunehmend bedeutende Rolle spielen Datensicherheit und Datenschutz. So sind Lösungen erforderlich, die eine verlässliche Transparenz in Bezug auf die Sicherheit der App und den Umgang der App mit den Daten vermitteln. Gerade für die Nutzung der weit verbreiteten Smartphones auf Basis der wenig datenschutzfreundlichen technischen Ökosysteme von Apple, Google und Microsoft fehlt es noch an Konzepten zum

Umgang mit den hier entstehenden spezifischen Risiken. Insofern besteht großer und dringlicher Handlungsbedarf, technisch innovative Lösungen zur Abschottung und Absicherung von Apps gegenüber der Betriebssysteminfrastruktur zu finden. Gleichzeitig muss aber auch eine sorgfältige Abwägung der enormen, ggf. auch für die individuell betroffenen Patienten relevanten, Chancen mit den genannten Risiken stattfinden.

Unabhängig von den softwareseitig umgesetzten Maßnahmen, die durch das ResearchKit von Apple deutlich vereinfacht werden [42], liegt die Verantwortung für den ethisch und methodisch korrekten Umgang mit den anvertrauten Daten bei den App-Entwicklern und dem Studienteam. Um das Potential persönlicher Gesundheitsdaten, die über mobile Devices erfasst werden, auszunutzen, ist ein bewusster Umgang mit den Möglichkeiten und Grenzen entsprechender Werkzeuge erforderlich und im Studiendesign zu berücksichtigen. Bei zunehmender Anzahl von Forschungsprojekten, an denen sich ein Patient mittels persönlicher Erfassung seiner Gesundheitsdaten beteiligen kann, werden sich die Rolle des Studienteilnehmers und dessen Ansprüche zunehmend weiterentwickeln. Der Patient sollte daher zu jeder Zeit einen transparenten Überblick und die Kontrolle über die Verwendung sensibler, medizinischer und persönlicher Daten haben. Der in der Fachliteratur häufig verwendete Begriff „data donation“ bindet die Bereitstellung von Daten an die Erwartung, dass diese einem guten Zweck dienen und einen Mehrwert bringen. Entsprechende Informationen über Forschungserfolge könnten zukünftig Anreiz für Patienten sein, ihre Daten bereitzustellen.

Aktuell gibt es in Deutschland eine erste Studie, die auf dem ResearchKit von Apple aufbaut: Die Studie „Back on Track“<sup>13</sup> möchte die Therapieoptionen und deren Vor- und Nachteile für verschiedene Patientengruppen untersuchen. Wenn es um die Erforschung vergleichbar sensibler Forschungsthemen geht, wird das Vertrauen in die Forschungseinrichtung zukünftig vermutlich eine zunehmende Rolle spielen. Die Sicherstellung der Authentizität des App-Betreibers und damit der Datenschutz- und Datensicherheitsverantwortung, für die er bürgt, liegt jedoch außerhalb des Einflussbereiches von Apple. Entsprechend sollten Maßnahmen etabliert werden, die Patienten helfen, die Vertrauenswürdigkeit von App-Betreibern einschätzen zu können.

Die großen Smartphone-Hersteller und Anbieter der entsprechenden Software-Ökosysteme arbeiten daran, ihre Systeme auch für den beruflichen Einsatz und die daran geknüpften höheren Erwartungen an Vertraulichkeit und Datenschutz weiterzuentwickeln [43]. Ob diese Entwicklung auch dazu führt, dass künftig Forschungs-Apps sicherer als bisher im Rahmen dieser Ökosysteme angeboten und genutzt werden können,

---

13 <https://itunes.apple.com/de/app/back-on-track-kreuzbandriss/id1016535765?mt=8>

bleibt abzuwarten. Die Andersartigkeit von Digital-Health-Anwendungen gegenüber etablierten Angebotsklassen stellt nach [44] die entscheidende Hürde für den Transfer von Digital-Health-Anwendungen in den Versorgungs- und damit auch in den Forschungsalltag dar.

Die Entwicklung von Konzepten und Standards für den Umgang mit der patientenorientierten bzw. sensorgesteuerten Datenerfassung im Forschungskontext ist auch für weitere Bereiche von Bedeutung. So müssen unter anderem Fragen zum Verfahren mit sich widersprechenden Daten von Patient und Arzt, zur Bearbeitung oder Änderung von Patientenangaben durch den Patienten selbst oder durch den Arzt sowie zum Qualitätsmanagement der mobil erfassten Daten beantwortet werden. Insgesamt hat die qualitative wie quantitative Datenerhebung durch mobile Messung im Patienten, am Patienten oder in der Umgebung des Patienten derart stark zugenommen, dass das Segment der Medical Devices in der amerikanischen Medizininformatik mittlerweile die gleiche Bedeutung eingenommen hat wie das Segment der Medizintechnik. Diese Entwicklung verdeutlicht den Bedarf an Methoden, um die zunehmende Informationsflut von Daten, z. B. durch Echtzeitmessungen, sinnvoll verwerten und mit anderen Datenquellen in Bezug setzen zu können.

## Literatur

1. Zundel, K.M., *Telemedicine: history, applications, and impact on librarianship*. Bulletin of the Medical Library Association, 1996. **84**(1): S. 71 - 79.
2. Bietz, M.J., Bloss, C.S., Calvert, S. et al., *Opportunities and challenges in the use of personal health data for health research*. J Am Med Inform Assoc, 2015.
3. Schmidt, S., Rienhoff, O., *Interdisciplinary assessment of personal health monitoring*. 2013, IOS Press, Amsterdam.
4. IHHI *Patient Adoption of mHealth. Use, Evidence and Remaining Barriers to Mainstream Acceptance*. 2015. IMS Institute For Healthcare Informatics, [http://www.vcbeat.net/wp-content/uploads/2015/10/IHHI\\_Patient\\_Adoption\\_of\\_mHealth.pdf](http://www.vcbeat.net/wp-content/uploads/2015/10/IHHI_Patient_Adoption_of_mHealth.pdf) (Abruf: 2016-11-22).
5. Heinze, O. *Entwicklung einer iOS App zur Erfassung eines standardisierten elektronischen Notfalleinsatzprotokolls*. *Telemed*. 2012. Berlin, <http://www.telemed-berlin.de/archiv/telemed2012/programm-2012-download> (Abruf: 2016-02-05).
6. Ebner-Priemer, U.W., Bohus, M., *Elektronische Tagebücher in der Medizin: Echtzeiterfassung von Symptomen*. Deutsches Ärzteblatt, 2008. **105**(28-29): S. 1550 - 1552.

7. Wilhelm, P., Perrez, M., *Ambulantes Assessment in der klinischen Psychologie und Psychiatrie*. Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 2008. **56**(3): S. 169 - 179.
8. Chang, C.-H., *Patient-reported outcomes measurement and management with innovative methodologies and technologies*. Quality of Life Research, 2007. **16**(1): S. 157 - 166.
9. FDA *Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. 2009. U.S. Food and Drug Administration, <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf> (Abruf: 2016-11-18).
10. Banos, O., Villalonga, C., Damas, M., Gloesekoetter, P., Pomares, H., Rojas, I., *PhysioDroid: Combining Wearable Health Sensors and Mobile Devices for a Ubiquitous, Continuous, and Personal Monitoring*. TheScientificWorldJournal, 2014. **2014**: S. 11 pages.
11. EC *Grünbuch über Mobile-Health-Dienste („mHealth“)*. 2014. Europäische Kommission, <http://ec.europa.eu/digital-agenda/en/news/green-paper-mobile-health-mhealth> (Abruf: 2016-11-22).
12. Wegscheider, K., Koch-Gromus, U., *Die Versorgungsforschung als möglicher Profiteur von Big Data*. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, 2015. **58**(8): S. 806-812.
13. Jardine, J., Fisher, J., Carrick, B., *Apple's ResearchKit: smart data collection for the smartphone era?* J R Soc Med, 2015. **108**(8): S. 294-296.
14. Langemak, S., *Apples ResearchKit: Der Retter kommt mit einer Reihe von Problemen*. Medscape, 2015. **18.03.2015**, <http://praxis.medscapemedizin.de/artikelansicht/4903430> (Abruf: 2016-11-22).
15. Ritter, S., *Apple's Research Kit Development Framework for iPhone Apps Enables Innovative Approaches to Medical Research Data Collection*. J Clin Trials, 2015. **5**(2): S. e120.
16. EMA *ICH Topic E 6 (R1) - Guideline for Good Clinical Practice* 2002. European Medicines Agency, [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500002874.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf) (Abruf: 2016-11-18).
17. FDA *CFR - Code of Federal Regulations Title 21 Part 11*. 2014: U.S. Food and Drug Administration, <http://www.ecfr.gov/cgi-bin/text-idx?SID=23dddc1a5b346f816d53acabd66fe3e2&mc=true&node=pt21.1.11&rgn=div5> (Abruf: 2016-11-22).
18. FDA *Guidance for Industry. Computerized Systems Used in Clinical Investigations*. 2007. U.S. Food and Drug Administration, <http://www.fda.gov/OHRMS/DOCKETS/98fr/04d-0440-gdl0002.pdf> (Abruf: 2016-11-18).

19. COE Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. 2007. The Council of the European Communities, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0042:20071011:en:PDF> (Abruf: 2016-11-22).
20. Albrecht, U.-V., *Biomedical Research Goes Mobile: App-Baukästen für die Forschung. Themen der Zeit: Kommentar*. Deutsches Ärzteblatt, 2015. **112**(17): S. A-774.
21. EC Medical Devices: *Guidance document - Qualification and Classification of stand alone software. Guidelines on the qualification and classification of stand alone software used in healthcare within the regulatory framework of medical devices*. 2016. European Commission, MEDDEV 2.1/6, <http://ec.europa.eu/DocsRoom/documents/17921/attachments/1/translations/en/renditions/pdf> (Abruf: 2016-11-24).
22. Albrecht, U.-V., Hrsg. *Chancen und Risiken von Gesundheits-Apps (CHARISMHA)*. 2016, Medizinische Hochschule, Hannover, <http://www.digibib.tu-bs.de/?docid=00060000> (Abruf: 2016-11-22).
23. Gietzelt, M., Wolf, K.-H., Haux, R., *A Nomenclature for the Analysis of Continuous Sensor and Other Data in the Context of Health-Enabling Technologies*, in *User Centered Networked Health Care. Proceedings of MIE 2011*, Hrsg.: A. Moen, et al. 2011, IOSPress, Amsterdam. S. 460-464.
24. Marschollek, M., Gietzelt, M., Schulze, M., Kohlmann, M., Song, B., Wolf, K.-H., *Wearable Sensors in Healthcare and Sensor-Enhanced Health Information Systems: All Our Tomorrows?* Healthc Inform Res, 2012. **18**(2): S. 97-104.
25. Knöppler, K., Neisecke, T., Nölke, L. *Digital-Health-Anwendungen für Bürger. Kontext, Typologie und Relevanz aus Public-Health-Perspektive. Entwicklung und Erprobung eines Klassifikationsverfahrens*. 2016. Bertelsmann Stiftung, [https://www.bertelsmann-stiftung.de/fileadmin/files/BSSt/Publikationen/GrauePublikationen/Studie\\_VV\\_Digital-Health-Anwendungen\\_2016.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSSt/Publikationen/GrauePublikationen/Studie_VV_Digital-Health-Anwendungen_2016.pdf) (Abruf: 2016-11-22).
26. Lucht, M., Bredenkamp, R., Boeker, M., Kramer, U. *Gesundheits- und Versorgungs-Apps. Hintergründe zu deren Entwicklung und Einsatz*. 2015. Universitätsklinikum Freiburg, Studienzentrum im Auftrag der Techniker Krankenkasse, <https://www.tk.de/centaurus/servlet/contentblob/724464/Datei/143235/Studie-Gesundheits-und-Versorgungs-Apps.pdf> (Abruf: 2016-11-22).

27. Gigerenzer, G., Schlegel-Matthies, K., Wagner, G.G. *Digitale Welt und Gesundheit. eHealth und mHealth – Chancen und Risiken der Digitalisierung im Gesundheitsbereich*. 2016. Sachverständigenrat für Verbraucherfragen beim Bundesministerium der Justiz und für Verbraucherschutz <http://www.svr-verbraucherfragen.de/wp-content/uploads/2016/01/Digitale-Welt-und-Gesundheit.pdf> (Abruf: 2016-11-22).
28. Gwaltney, C.J., Shields, A.L., Shiffman, S., *Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures: A Meta-Analytic Review*. *Value in Health*, 2008. **11**(2): S. 322 - 333.
29. Stone, A.A., Shiffman, S., Schwartz, J.E., Broderick, J.E., Hufford, M.R., *Patient non-compliance with paper diaries*. *BMJ*, 2002. **2002**(324): S. 1193 - 1194.
30. Coons, S.J., Gwaltney, C.J., Hays, R.D. et al., *Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report*. *Value in Health*, 2009. **12**(4): S. 419 - 429.
31. Wilcox, A.B., Gallagher, K.D., Boden-Albala, B., Bakken, S.R., *Research Data Collection Methods: From Paper to Tablet Computers*. *Medical Care*, 2012. **50**: S. S68 - S73.
32. Rose, M., Bezjak, A., *Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples*. *Quality of Life Research*, 2009. **18**(1): S. 125 - 136.
33. Becker, S., Miron-Shatz, T., Schumacher, N., Krocza, J., Diamantidis, C., Albrecht, U.-V., *mHealth 2.0: Experiences, Possibilities, and Perspectives*. *JMIR mHealth and uHealth*, 2014. **2**(2): S. e24.
34. Duhaime-Ross, A., *Apple's new ResearchKit: ‚Ethics quagmire‘ or medical research aid? The software platform could be great for medicine, but it has some growing up to do*. *The Verge*, 2015 (e-Pub: 10.03.2015), <http://www.theverge.com/2015/3/10/8177683/apple-research-kit-app-ethics-medical-research> (Abruf: 2016-11-22).
35. Fahrenberg, J., *Ambulantes Assessment*, in *Handbuch Statistik, Methoden und Evaluation*, Hrsg.: H. Holling and B. Schmitz. 2010, Hogrefe, Göttingen. S. 201-212.
36. Pommerening, K. *Personalisierte Medizin und Informationstechnik - Aspekte des Datenschutzes*. 55. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS). 2010. Mannheim, <http://www.gmds2010.de/wp-content/uploads/fohlen/91.pdf> (Abruf: 2016-11-22).
37. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.

38. EC *Draft Code of Conduct on privacy for mobile health applications*. 2016. European Commission, [http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc\\_id=16125](http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=16125) (Abruf: 2016-11-22).
39. Neumann, K., Larisch, K., Dietzel, J., Kurepkat, M., Weißer, M., Wenzlau, V. *Digitale Versorgungsprodukte. Chancen nutzen, sichere Wege in den Markt schaffen*. 2016. IGES, Studienbericht im Auftrag der Techniker Krankenkasse, <https://www.tk.de/centaurus/servlet/contentblob/915490/Datei/175314/TK-Pressemappe-IGES-Studie-Digitale-Angebote-in-der-Gesundheitsversorgung.pdf> (Abruf: 2016-11-22).
40. Fahrenberg, J., Leonhart, R., Foerster, F., Hrsg. *Alltagsnahe Psychologie. Datenerhebung im Feld mit hand-held PC und physiologischem Mess-System*. 2011, e-Buch, [http://www.jochen-fahrenberg.de/uploads/media/ALLTAGSNAHE\\_PSYCHOLOGIE.pdf](http://www.jochen-fahrenberg.de/uploads/media/ALLTAGSNAHE_PSYCHOLOGIE.pdf) (Abruf: 2013-04-19).
41. Kroll, R.R., Boyd, J.G., Maslove, D.M., *Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study*. J Med Internet Res, 2016. **18**(9): S. e253.
42. Stöbe, M., *Probanden-Finder. Wie Apples ResearchKit funktioniert und die Forschung erleichtert*. c't, 2016. **2016**(3): S. 174-179.
43. Weber, V., *Gemeinsam getrennt. Smartphones und Tablets beruflich nutzen*. c't, 2016. **2016**(3): S. 170-172.
44. Knöppler, K., Oschmann, L., Neumann, J., Neisecke, T. *Transfer von Digital-Health-Anwendungen in den Versorgungsalltag. Teil 1: Transfermodell, Varianten und Hürden*. 2016. Bertelsmann Stiftung, Teilbericht, vorläufige Fassung, August 2016, [https://www.bertelsmann-stiftung.de/fileadmin/files/BSSt/Publikationen/GrauePublikationen/Studie\\_VV\\_Transfer-von-DH-Anwendungen\\_Teil1\\_2016.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSSt/Publikationen/GrauePublikationen/Studie_VV_Transfer-von-DH-Anwendungen_Teil1_2016.pdf) (Abruf: 2016-11-22).

# Sekundärnutzung klinischer Daten

*Johannes Drepper, Hans-Ulrich Prokosch, Martin Dugas, Sebastian C. Semler*

## Anwendungsbereich

Eine Reihe von Projekten auf nationaler [z. B. 1, 2] und internationaler Ebene [z. B. 3] haben sich in den letzten Jahren mit dem Thema auseinandergesetzt, wie die im Behandlungsfall dokumentierten medizinischen Daten in Krankenhäusern (und ggf. auch im ambulanten Bereich) für Forschungs- oder Qualitätssicherungsmaßnahmen sekundär genutzt werden können. Parallel zu und z. T. verzahnt mit dieser Projektaktivität wurde mit Unterstützung der TMF die Arbeitsgruppe „Nutzung von elektronischen Patientenakten für die klinische Forschung“ in der GMDS gegründet und etabliert, die sich seit 2009 ebenfalls explizit dem Thema der Sekundärnutzung widmet.

Dieser schon länger feststellbare Trend zur vermehrten Sekundärnutzung klinischer Behandlungsdaten [vgl. 4] stützt sich auf eine Reihe relevanter Anwendungsfälle. Diese reichen von Machbarkeitsstudien [5] über die Rekrutierungsunterstützung [6] bis hin zur gleichzeitigen Verwendung klinischer Daten sowohl für Forschung als auch Versorgung nach dem Single-Source-Paradigma [7]. Zudem basieren epidemiologische Studien und Auswertungen in der Versorgungsforschung immer häufiger auf bereits in der Versorgung erhobenen Daten [vgl. 8]. Eine vermehrt untersuchte und dabei hoch relevante Fragestellung ist z. B. die nach bisher unentdeckten Nebenwirkungen von Medikamenten, die erst nach der Zulassung und bei entsprechend breiter Anwendung mit Hilfe großer Datensätze gefunden werden können [9].

Ausgenommen sind in diesem Kapitel Fragestellungen zur Nutzung der rein zu Abrechnungs- oder Qualitätssicherungszwecken von den behandelnden Einrichtungen erhobenen und an externe Stellen übertragenen Daten. Auch wenn diese Art der Datennutzung, oft auch als „Routinedatennutzung“ bezeichnet, ebenfalls mit unter den Oberbegriff der Sekundärdatennutzung fällt [8], so spricht doch viel dafür, diese beiden Varianten der Sekundärdatennutzung hinsichtlich der notwendigen IT-Unterstützung getrennt zu behandeln. Da jedoch diese Datenbestände und ihre Nutzbarkeit, nicht zuletzt auch durch die Regelungen zur Datentransparenz im Sozialgesetzbuch V (§§ 303a-e SGB V) mit den darin festgelegten Nutzungsmöglichkeiten für Forschungseinrichtungen und die hierzu von der TMF im Rahmen des Forums Versorgungsforschung

durchgeführten Veranstaltungen<sup>12</sup>, zunehmend auch von der akademischen Forschung in den Blick genommen werden, sollte auch dieser Bereich in einem der kommenden Berichte ausführlich dargestellt werden.

Ebenfalls ausgenommen ist die Sekundärnutzung von Daten aus klinischen Studien. Hinweise auf entsprechende Modelle, Plattformen und Nutzungsarten finden sich im Kapitel zu klinischen Studien in diesem Buch.

Zu den im Folgenden betrachteten Anwendungsfällen:

In Machbarkeitsstudien (Feasibility Studies) wird überprüft, ob in den behandelnden Einrichtungen künftig mit ausreichend Patienten für ein bestimmtes klinisches Forschungsvorhaben gerechnet werden kann. Hierzu wird anhand der Daten der Behandlungsfälle aus der Vergangenheit überprüft, wieviele Patienten mit den relevanten Ein- und Ausschlusskriterien behandelt wurden. Diese Zahlen dienen als Grundlage für eine bestmögliche Schätzung künftig ansprechbarer Patienten für ein entsprechendes Forschungsvorhaben. Da die Daten auf Behandlungsfällen aus der Vergangenheit basieren, gehören die von einer solchen Datennutzung betroffenen Datensätze im Regelfall nicht zu Patienten, die für eine kommende Rekrutierung direkt in Frage kommen. Insofern können die Auswertungen auch auf anonymisierten Datenbeständen ausgeführt werden.

Eine anonymisierte Datenverarbeitung verbietet sich hingegen bei dem Anwendungsfall der Rekrutierungsunterstützung, da hier tatsächlich die Datensätze der Patienten gesucht werden, die konkret für ein Forschungsvorhaben angesprochen werden können. In diesem Fall kann die Auswertung zwar pseudonym erfolgen, die Ansprache der Patienten durch die in die Behandlung eingebundenen Personen erfolgt aber auf Basis der direkt identifizierenden Daten der Patienten.

Noch wenig verbreitet ist die unmittelbare bzw. zumindest halbautomatische Übernahme klinischer Daten, die im Versorgungskontext dokumentiert wurden, in Electronic Data Capture (EDC) Systeme für klinische Studien. Es ist zu erwarten, dass in der Zukunft dieses Verfahren an Bedeutung gewinnen wird, da die elektronische Patientenakte mehr und mehr Quelldaten (Source Data) für klinische Studien enthalten wird. Bereits heute ist eine Reihe von klinischen Daten ausschließlich in der elektronischen Patientenakte verfügbar. Diese Daten werden direkt in einer Form und Qualität erhoben und elektro-

---

1 Siehe <http://www.tmf-ev.de/news/1464>.

2 Die TMF hat in Kooperation mit dem DIMDI als Begleitstruktur zum Informationssystem Versorgungsdaten des DIMDI den Arbeitskreis Versorgungsdaten eingerichtet: [http://www.tmf-ev.de/Arbeitsgruppen\\_Foren/ForumVersorgungsforschung/ArbeitskreisVersorgungsdaten.aspx](http://www.tmf-ev.de/Arbeitsgruppen_Foren/ForumVersorgungsforschung/ArbeitskreisVersorgungsdaten.aspx).

nisch erfasst, so dass sie sowohl als Dokumentation des Versorgungsprozesses als auch direkt für klinische Studien mitgenutzt werden können [7, 10]. Ein alternativer Workflow sieht eine Vorausfüllung der elektronischen Studienformulare (eCRFs) durch die in der Behandlung dokumentierten und elektronisch erfassten Daten vor [11, 12]. Für eine derartige Nutzung wurden innerhalb des europäischen EHR4CR-Projekts anhand 23 ausgewählter klinischer Studien die 131 am häufigsten genutzten Datenelemente identifiziert und deren Verfügbarkeit in den KAS-Datenbanken der Kliniken überprüft [13]. Eine Herausforderung ist hierbei, die Erfüllung der Kriterien hinsichtlich Qualitätssicherung und Validierung gemäß der internationalen Richtlinie Good Clinical Practice (GCP) für die diversen klinischen Subsysteme sicherzustellen, welche diese Quelldaten liefern.

Ein weiteres Anwendungsfeld für die Sekundärnutzung klinischer Daten entsteht durch den systematischen Aufbau von Biobanken an den klinischen Standorten in Deutschland.<sup>3</sup> Die klinische Annotation der im Versorgungskontext erhobenen Proben kann später für eine Vielzahl von Forschungsprojekten genutzt werden, die z. T. auch Therapieoptionen zum Ziel haben, die heute mit dem Schlagwort der „Präzisionsmedizin“ verbunden werden. Im Regelfall werden dann durch die behandelnden Einrichtungen präventiv weitreichende Einwilligungen der Patienten für die spätere Sekundärnutzung der Proben und zugehöriger klinischer Daten eingeholt [14, 15].

## Betreiber, Anwender, Nutzer

Im Unterschied zur rein forschungsspezifischen IT sind bei der Sekundärnutzung klinischer Daten auch die IT-Systeme der Versorgung mit ihren jeweiligen Anwendern und Nutzern zu berücksichtigen. Als Betreiber der Versorgungs-IT sind im Regelfall lokale und der behandelnden Einrichtung zugehörige Abteilungen oder auch „Rechenzentren“ anzusehen. Die Anwendung der einzelnen Softwarekomponenten eines KIS im Sinne eines Customizings erfolgt ebenfalls lokal, was häufig zu unterschiedlichen Dokumentationskonzepten führt, auch wenn dieselbe Grundsoftware eingesetzt wird. Als Nutzer sind schließlich alle in die Dokumentation im Rahmen der Versorgung eingebundenen Personen anzusehen.

An einigen universitären Standorten in Deutschland kooperieren die Universitätsklinika an der Schnittstelle zwischen Versorgung und Forschung auch eng mit außeruniversitären oder außerklinischen Einrichtungen, um so einen besseren Know-how-Transfer zwischen Versorgung und Forschung zu ermöglichen. Allerdings steht bei diesen

---

3 Vgl. hierzu auch das Kapitel zu Biobanken im vorliegenden Bericht.

Kooperationen immer auch die gemeinsame Nutzung von Datenverarbeitungs- oder Speicherressourcen der außerklinischen Partner auf der Agenda. Prominente Beispiele solcher Kooperationen sind das Nationale Centrum für Tumorerkrankungen (NCT) in Heidelberg, welches als gemeinsame Einrichtung des Universitätsklinikums und des DKFZ (mitgetragen von der Deutschen Krebshilfe) arbeitet, sowie das Berlin Institute of Health (BIH) als Zusammenschluss der Charité und des Max-Delbrück-Centrums für Molekulare Medizin. Weitere Beispiele einer solchen mehr oder weniger institutionalisierten Zusammenarbeit sind bekannt.

Die Forschungs-IT ist hingegen schon seit vielen Jahren von einem Trend hin zur Verbundforschung und einer entsprechenden einrichtungsübergreifenden Kooperation geprägt. Im Rahmen kommerzieller Studien sind regelmäßig pharmazeutische Unternehmen oder CROs als Betreiber anzusprechen. Aber auch in der akademischen Forschung findet diesbezüglich ein Zentralisierungsprozess statt, so dass auch in multizentrischen IITs häufig ein zentral betriebenes EDC-System eingesetzt wird. Auch durch den Aufbau der Deutschen Zentren der Gesundheitsforschung wird die Planung und Implementierung der in der Forschung genutzten IT-Systeme weiter zentralisiert.

Diese Zentralisierung betrifft ebenfalls die administrative Anwendung der Forschungs-IT und entsprechend das Customizing der Softwaresysteme. Im Regelfall liegen der Betrieb – auch im Sinne eines Hostings – und die administrative Anwendung der Systeme in einer Hand, auch wenn ein Trend hin zu einem vermehrten extern genutzten Hosting der Softwaresysteme geht.<sup>4</sup> Die bei einer vermehrten – auch internationalen – Zentralisierung der Verarbeitung sensibler Gesundheitsdaten zu berücksichtigenden rechtlichen Rahmenbedingungen werden im Kapitel zum Cloud-Computing im vorliegenden Report ausführlicher beleuchtet. Die Nutzer der Forschungs-IT in den Krankenhäusern sind z. T. dieselben wie im Bereich der Versorgungs-IT. Daneben gibt es auch noch speziell für die Forschung eingestelltes Personal wie z. B. Study Nurses.

Bei der Betrachtung des von der Sekundärnutzung klinischer Daten betroffenen Personenkreises sind somit in den Bereichen der Forschung und Versorgung jeweils unterschiedliche Organisationsprinzipien zu berücksichtigen. Entsprechend sind für die strategischen Vorgaben in den beiden Bereichen auch unterschiedliche Personenkreise verantwortlich, was sich in Bezug auf die notwendige Harmonisierung von Schnittstellen, Standards und Dokumentationskonzepten als hohe Hürde erweist.

---

4 Vgl. das Kapitel zu klinischen Studien im vorliegenden Bericht.

## Rahmenbedingungen

Auch die rechtlichen, organisatorischen und technischen Rahmenbedingungen der Sekundärnutzung klinischer Daten in der Forschung sind davon gekennzeichnet, dass hinsichtlich Versorgung und Forschung zwei ganz unterschiedliche Sphären mit ihren je eigenen Spielregeln zu betrachten sind.

### Rechtliche Rahmenbedingungen

Klinische Daten, die personenbezogen oder personenbeziehbar sind, unterliegen grundsätzlich datenschutzrechtlichen Bestimmungen und sind diesen zufolge, da sie die Gesundheit betreffen, zudem als besonders schützenswert eingestuft. Für diese Daten gilt somit immer eine strenge Zweckbindung und eine Zweckänderung, wie sie die Sekundärnutzung per Definition darstellt, erfordert eine eigenständige rechtliche Grundlage. Das Datenschutzrecht sieht im Regelfall die Einholung einer informierten und freiwilligen schriftlichen Einwilligungserklärung als Grundlage einer rechtskonformen Verarbeitung zu sekundären Zwecken vor. Zudem ist zu beachten, dass die im Behandlungskontext erhobenen Daten im Regelfall der ärztlichen Schweigepflicht unterliegen, die als eigenständige zweite „Schranke“ vor einer Herausgabe der Daten anzusehen ist. Insofern muss die datenschutzrechtliche Einwilligungserklärung auch den Kriterien einer Schweigepflichtsentbindung entsprechen.

In vielen Fällen der Sekundärnutzung klinischer Daten ist allerdings eine ausführlichere Betrachtung der gesetzlichen Rahmenbedingungen notwendig. Diese sind leider unübersichtlicher als oft dargestellt. Zunächst ist zu prüfen, ob es in dem betreffenden Bundesland ein Landeskrankenhausgesetz oder eine vergleichbar spezifische gesetzliche Regelung gibt, und ob diese Regelung auf das jeweilige Krankenhaus anwendbar ist. Ist dies der Fall und enthält das betreffende Gesetz auch Regelungen zum Datenschutz und zur Sekundärnutzung, so sind diese auch anzuwenden. Nachgeordnet ist im Regelfall das jeweilige Landesdatenschutzrecht anzuwenden, wobei berücksichtigt werden muss, dass viele Landesdatenschutzgesetze für öffentliche Einrichtungen der Länder, die im Wettbewerb mit privatwirtschaftlichen Einrichtungen stehen, auch auf unterschiedliche Regelungsbereiche des Bundesdatenschutzgesetzes (BDSG) verweisen.<sup>5</sup>

Wenn Daten oder Proben aufgrund einer Einwilligungserklärung der Patienten sekundär für die Forschung genutzt werden sollen, so muss diese Erklärung bestimmt sein, so

---

5 Weitere Ausnahmen von der dargestellten Situation bestehen für viele Krankenhäuser in kirchlicher Trägerschaft und auch Krankenhäuser des Bundes (Bundeswehrkrankenhäuser). Weitere Hinweise hierzu finden sich ebenfalls in der Darstellung von Schneider [16].

dass klar zu erkennen ist, unter welchen Bedingungen sich die betroffene Person mit der Erhebung, Verarbeitung oder Nutzung welcher Daten einverstanden erklärt. 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 wie etwa Gesundheitsdaten bezieht. Der Aufbau großer klinischer Biobanken beispielsweise geht jedoch regelmäßig mit der Hoffnung einher, diese Daten später auch für neue und ggf. heute noch nicht absehbare Forschungsfragestellungen nutzen zu können. Der Arbeitskreis Medizinischer Ethik-Kommissionen in Deutschland hat hierzu Musterformulierungen erarbeitet, die sowohl der informationellen Selbstbestimmung der Patienten gerecht werden, als auch eine möglichst zweckoffene und ggf. auch krankheitsübergreifende Nutzung von Proben im Kontext der medizinischen Forschung ermöglichen [17]. Zudem erläutern die generischen Datenschutzkonzepte der TMF, mit welchen technischen und organisatorischen Maßnahmen die durch eine langfristige und vergleichsweise zweckoffene Nutzung von Proben und Daten entstehenden Risiken in Bezug auf die informationelle Selbstbestimmung ausbalanciert werden können [18].

Darüber hinaus enthalten jedoch einige Landeskrankenhausgesetze zusätzliche Anforderungen an die Einwilligungserklärung von Patienten, wenn sich diese auf im Rahmen der Versorgung erhobene Daten bezieht. Dabei ist insbesondere festzustellen, dass die Voraussetzungen der Bundesländer sehr unterschiedlich sind. In einigen Landeskrankenhausgesetzen ist eine Einwilligung in die Nutzung oder auch Übermittlung der Patientendaten zu sekundären Zwecken nur „im Einzelfall“ oder mit einem konkreten Vorhabensbezug möglich. Insofern ist festzustellen, dass für solche Einwilligungsszenarien, wenn sie bundeslandübergreifend umgesetzt werden sollen und gleichzeitig Forschungszwecke nur wenig eingeschränkt werden können, noch mit erheblichen Rechtsunsicherheiten behaftet sind. Dies betrifft auch die im Rahmen des Förderkonzepts Medizininformatik des BMBF umzusetzenden Datenintegrationszentren an den Universitätsklinika und ggf. Partnerstandorten.<sup>6</sup>

Nun gibt es aber eine Reihe von Anwendungsfällen, die zu dem hier skizzierten Regelfall aus unterschiedlichen Gründen nicht gut passen, insbesondere weil z. B. die Einholung einer Einwilligung aller Betroffenen kaum oder gar nicht möglich ist. Für solche Fälle sind die sogenannten Forschungsklauseln im Datenschutzrecht gedacht, die unter bestimmten Umständen und gemäß einer Abwägung der relevanten Grundrechte – der informationellen Selbstbestimmung auf der einen und der Forschungsfreiheit auf der anderen Seite – auch eine Nutzung von Daten für die Forschung ohne Einwilligung erlauben. Allerdings sind gerade diese Forschungsklauseln in den verschiedenen Da-

---

6 Siehe <http://www.gesundheitsforschung-bmbf.de/de/medizininformatik.php>.

tenschutzgesetzen nicht gleichermaßen enthalten bzw. auch nicht einheitlich formuliert. Verkomplizierend kommt hinzu, dass z. T. ähnliche Klauseln auch in manchen Landeskrankenhausgesetzen formuliert sind und diese, wenn anwendbar, vorrangig zu berücksichtigen sind. Allerdings sind auch diese Regelungen von Bundesland zu Bundesland sehr unterschiedlich.

Solche spezialgesetzlichen Regelungen für den Krankenhausbereich können auch Offenbarungsbefugnisse im Sinne der Regelungen zur ärztlichen Schweigepflicht enthalten. Alleine im Datenschutzrecht verankerte Regelungen zur Sekundärnutzung stellen hingegen keine vergleichbaren Offenbarungsbefugnisse dar. Somit können spezialgesetzliche Regelungen in den Landeskrankenhausgesetzen in bestimmten Fällen auch die Herausgabe personenbezogener Daten ohne explizite Einwilligung und Schweigepflichtsentbindung der Betroffenen rechtskonform ermöglichen, wohingegen rein datenschutzrechtliche Erlaubnistatbestände auf die sekundäre Nutzung der Daten in der jeweiligen behandelnden Einrichtung bzw. Abteilung beschränkt bleiben.

Die TMF hat diese Rechtslage in 2014 im Rahmen eines Gutachtens für das Projekt cloud4health<sup>7</sup> umfassend und detailliert für alle Krankenhausarten, alle Bundesländer und verschiedene Anwendungsszenarien der Sekundärnutzung aufarbeiten lassen, so dass hier auf diese übersichtliche Darstellung zu der komplexen Sachlage verwiesen werden kann [16].<sup>8</sup>

Jenseits der Frage, welche datenschutzrechtlichen Regelungen in einem spezifischen Fall anzuwenden sind, ist das vorliegende Rechtsgutachten von Schneider auch auf einen für die Sekundärnutzung relevanten Aspekt eingegangen, der eigentlich zu den mehr oder weniger einheitlich geregelten Grundzügen des Datenschutzes gehört. Dabei geht es um die Frage, ob pseudonymisierte Daten bei einer Stelle, die selbst nicht über den Zuordnungsschlüssel zur Auflösung der Pseudonymisierung verfügt, als personenbeziehbar anzusehen sind und ihre Nutzung durch diese Stelle entsprechend gemäß den Regeln des Datenschutzrechts auszugestalten ist. Diese Frage, die einen zentralen Aspekt vieler Szenarien zur Sekundärnutzung betrifft, wird leider bis heute uneinheitlich beantwortet. Hintergrund dieser bestehenden Unklarheit sind offensichtlich mehrdeutig interpretierbare Formulierungen in den entsprechenden Gesetzestexten. Da auch die Aufsichtsbehörden in Deutschland diesbezüglich keine einheitliche Meinung vertreten, können sich größere Verbundprojekte typischerweise nur auf den

---

7 Siehe [www.cloud4health.de](http://www.cloud4health.de).

8 Ergänzend zu dem Rechtsgutachten von Schneider bietet die TMF ein Online-Tool zur Recherche der relevanten Rechtsgrundlagen an. Als Rahmenbedingungen können das Bundesland, die Art des Krankenhauses und der Zweck der Datennutzung (Forschung oder Qualitätssicherung) angegeben werden. Siehe [www.tmf-ev.de/irene](http://www.tmf-ev.de/irene).

datenschutzrechtlich „kleinsten gemeinsamen Nenner“ verlassen: die Annahme eines absoluten Personenbezugs. Das bedeutet, dass pseudonymisierte Daten immer als personenbeziehbar anzusehen wären, auch wenn die verarbeitende Stelle selbst keinen Zugriff auf die der Pseudonymisierung zugrundeliegende Zuordnungsregel besitzt. Der Europäische Gerichtshof (EuGH) hat in einem aktuellen Urteil zur Personenbeziehbarkeit von IP-Adressen grundsätzlich eine relative Sicht der Personenbeziehbarkeit unterstützt [19]. Ob und wie dieses Urteil die Rechtspraxis in Deutschland beeinflusst, bleibt jedoch abzuwarten.

Das Konzept des absoluten Personenbezugs scheint für viele Szenarien der Sekundärnutzung eine hohe Hürde darzustellen, da in Folge der Anwendung dieses Konzepts regelmäßig die Einholung einer informierten und freiwilligen Einwilligung der betroffenen Patienten erforderlich sein dürfte. Letztlich ist aber heute zumeist davon auszugehen, dass die immer umfangreicheren medizinischen Daten selbst schon ein erhebliches Reidentifizierungspotential besitzen. Zudem ist zu berücksichtigen, dass im Zeitalter umfangreich öffentlich verfügbarer Informationen im Internet und insbesondere auch in sozialen Netzwerken wie etwa Facebook von einem immer größeren und kaum noch eingrenzbaren Zusatzwissen bei allen datenverarbeitenden Stellen auszugehen ist. Insofern ist die Frage der Relativität des Personenbezugs pseudonymisierter Daten zwar immer noch für ganz bestimmte Anwendungsfälle mit entsprechend eingeschränkten medizinischen Daten von hoher Relevanz. Für die allermeisten Anwendungsfälle mit etwas umfangreicheren Datensätzen wird man aber heute im Regelfall von einer Personenbeziehbarkeit ausgehen müssen, so dass in diesen Fällen immer eine entsprechende datenschutzrechtliche Grundlage für die Sekundärnutzung unverzichtbar ist.

Alternativ zur Pseudonymisierung kann auch eine Anonymisierung der klinischen Daten vor einer Sekundärnutzung in Betracht gezogen werden. Die fehlende Rückverfolgbarkeit der Daten schränkt dann allerdings, wie für die Rekrutierungsunterstützung bereits beschrieben, die möglichen Anwendungsfälle stark ein. Für eine wirksame Anonymisierung klinischer Daten ist zudem im Regelfall eine Vergröberung bzw. Veränderung der Daten erforderlich (k-Anonymisierung, l-Diversität), die die Verwendbarkeit für wissenschaftliche Fragestellungen weiter einschränkt [20]. Für eine ausführlichere Darstellung der Anforderungen an eine Anonymisierung und der sich daraus ergebenden Konsequenzen wird auf das separate Kapitel zur Anonymisierung in diesem Report verwiesen.

Im Frühjahr 2016 wurde nach vierjähriger Verhandlung die EU-Datenschutzgrundverordnung (EU-DSG-VO) verabschiedet, die ab Mai 2018 anwendbar sein wird. Damit wird das grundsätzliche Niveau des Datenschutzes in der EU neu definiert [21]. Ein

wesentliches Ziel dieser umfangreichen Gesetzgebungsinitiative war die weitergehende Harmonisierung des Datenschutzrechts in der EU. Ein solcher Harmonisierungseffekt wäre gerade für die heterogene Rechtslage zum Datenschutz für die biomedizinische Forschung in Deutschland sehr wünschenswert. Allerdings lässt die EU-DSG-VO den Nationalstaaten in Art. 9 Abs. 4 gerade in Bezug auf Gesundheitsdaten viel Spielraum für eigene und ggf. auch engere Regelungen. Derzeit ist noch sehr offen, wie die Gesetzgebung im Bund und in den Ländern mit diesem Spielraum umgehen wird.

## **Technische und organisatorische Rahmenbedingungen**

Eine zentrale Voraussetzung für die Sekundärnutzung klinischer Daten ist zunächst, dass bekannt ist, welche Daten überhaupt erfasst werden und für die Forschung zur Verfügung stehen. Gerade bei einrichtungsübergreifenden Anwendungsfällen kann es zu erheblichen Aufwänden führen, herauszubekommen, welche Daten wo erhoben werden. Da die Struktur der erhobenen Daten bzw. die Metadaten nicht öffentlich zugänglich sind, ist es in vielen Fällen praktisch unmöglich, zu beurteilen, welche Datenelemente in verschiedenen Datenquellen verfügbar sind und daher für eine gemeinsame wissenschaftliche Auswertung in Frage kommen. Seitens der Wissenschaft wird daher freier Zugang zu den Metadaten aus Studien („Open Metadata“) gefordert [22, 23] und es werden Informationsinfrastrukturen benötigt, um diese Metadaten für die Scientific Community zur Verfügung zu stellen [24].<sup>9</sup>

Eine weitere wichtige Voraussetzung für die Nutzbarkeit klinischer Daten stellt deren ausreichende anwendungsfallbezogene Datenqualität dar. Die sekundäre Nutzung impliziert dabei immer, dass die Daten primär zu einem anderen Zweck erhoben und somit ggf. auch mit einer anderen Zielstellung qualitätsgesichert wurden. Nach wie vor werden in der klinischen Routine viele Daten zu Abrechnungszwecken oder auch zu Zertifizierungszwecken (z. B. als Onkologisches Zentrum) dokumentiert, so dass davon auszugehen ist, dass auch die qualitätssichernden Verfahren diese jeweiligen Ziele fokussieren. Dies führt sicher zu einer nahezu vollständigen Erfassung der direkt abrechnungsrelevanten Daten durch die Leistungserbringer und zudem zu einer gründlichen Prüfung durch die Kostenträger. Aber auch wenn die geplante Sekundärnutzung nur abrechnungsrelevante Daten wie etwa Diagnosen betrifft, kann nicht automatisch von einer ausreichenden Datenqualität ausgegangen werden, da z. B. die für eine wissenschaftliche Fragestellung relevanten Kriterien einer Diagnose andere sein können als im Abrechnungsfall. Bei nicht finanzwirksamen Nebendiagnosen muss zudem mit einer Unterkodierung gerechnet werden, oder im Falle der Berücksichtigung im Rah-

---

<sup>9</sup> Siehe hierzu auch das Kapitel zu Metadaten im vorliegenden Report..

men des morbiditätsorientierten Risikostrukturausgleichs auch mit einer kassengetriebenen Überkodierung.<sup>10</sup> Auch können sich Kodierungen aus Abrechnungsgründen im ambulanten und stationären Bereich unterscheiden. Letztlich kann nur die Empfehlung gegeben werden, in jedem Falle und mit Blick auf die konkreten Anforderungen des Anwendungsfalls die vorhandene Datenqualität kritisch zu prüfen [25].

Auch die in der Routinedokumentation verwendeten Kodierungs- und Klassifikationssysteme können für eine konkrete wissenschaftliche Nachnutzung ungeeignet sein. So können z. B. Patienten mit seltenen monogenetischen Erkrankungen oft anhand der Routinedaten nicht sicher identifiziert werden, da ihre Diagnosen im Klassifikationssystem der ICD- 10 in nicht differenzierenden „Sammel- oder Restklassen“ landen. Die Entwicklung hin zur individualisierten Medizin, mit ihrem molekularbiologischen Verständnis der Ätiologie, Pathogenese und Diagnose einer Krankheit, wird das Problem mit dem bestehenden Klassensystem der ICD- 10 noch weiter verschärfen [26]. Inwieweit das bei der WHO in Vorbereitung befindliche internationale Klassifikationssystem ICD-11 hierbei Verbesserungen bringen wird, bleibt abzuwarten.

Referenzterminologien wie z. B. SNOMED-CT könnten viele Anwendungsfälle sowohl im Versorgungs- als auch im Forschungskontext unterstützen und würden dabei einige der mit den heute verwendeten Kodiersystemen verbundenen Probleme beseitigen oder zumindest abschwächen [27, 28]. Insbesondere würde SNOMED-CT eine semantische Annotation von Patientenakten ermöglichen und wäre damit für eine integrative Analyse aus mehreren Datenquellen von großer Bedeutung [29, 30].

Damit könnte eine solche Referenzterminologie auch eine geeignete „Brücke“ für die Nutzung und korrekte Interpretation der nur einmal erfassten und kodierten Daten sowohl in der Versorgung als auch in der Forschung darstellen. Allerdings kommt eine Referenzterminologie wie SNOMED- CT in Deutschland derzeit sowohl in der Forschung als auch im Kontext der Routineversorgung insbesondere aus lizenzrechtlichen Gründen noch nicht zum Einsatz [31]. Allerdings wären bei einer Einführung von SNOMED- CT nicht nur die Lizenzkosten zu berücksichtigen. Die Kosten für den Aufbau einer Terminologie-Infrastruktur, die Aufwände bei Softwareherstellern und Anwendern sowie die Maintenance-Kosten dürften die Lizenzkosten deutlich übersteigen [28, 32]. Ohne einen zumindest schrittweisen Einstieg in die Nutzung von SNOMED- CT würde sich Deutschland jedoch immer weiter von der internationalen Entwicklung abkoppeln [30, 32, 33]. Konkret hat sich beispielsweise im europäischen Projekt epSOS zur grenzüberschreitenden Vernetzung der IT im Gesundheitswesen

---

<sup>10</sup> Siehe Interview mit dem Vorstandsvorsitzenden der Techniker Krankenkasse Jens Baas vom 9.10.2016 in der Frankfurter Allgemeinen Sonntagszeitung: <http://www.faz.net/-hbv-8m6up>.

gezeigt, dass bestimmte Entitäten nur auf der Basis von SNOMED-CT einheitlich codiert werden können [34].

Das Problem der Lizenzierung von SNOMED-CT sollte aber nicht den Blick auf andere, bereits heute kostenfrei oder günstig einsetzbare Kodiersysteme für Teilbereiche der medizinischen Dokumentation verstellen. So gibt es z. B. mit LOINC für den Bereich der Laboruntersuchungen ein mächtiges terminologisches System, welches zudem lizenzkostenfrei eingesetzt werden kann [35, 36]. Darüber hinaus wird in Kombination mit LOINC auch die Verwendung von UCUM für die standardisierte Erfassung der Messgrößen empfohlen [32, 37]. Zu den relevanten terminologischen und klassifikatorischen Standards in den Bereichen Pathologie (mit den dort angeschlossenen Biobanken) und Bildgebung sei auf die jeweiligen Kapitel im vorliegenden Report verwiesen. Zum intensivmedizinischen Bereich finden sich Hinweise in der Arbeit von Ingenerf und Ko-Autoren [38]. Auch wenn heute bei Weitem nicht für alle Bereiche der medizinischen Dokumentation lizenzkostenfreie und sowohl für die Forschung wie die Versorgung geeignete Standards existieren, zeigt das Beispiel LOINC doch, dass die Hürden für die Verwendung existierender Standards offensichtlich nicht rein lizenzrechtlicher oder kostentechnischer Natur sind.

Aufbauend auf einer ersten Pilotanalyse aus dem IT-Strategie-Projekt der TMF wurde im Rahmen des vom BMBF geförderten Projekts zur KIS-basierten Rekrutierungsunterstützung<sup>11</sup> untersucht, wieviele der in einer Zufallsauswahl von Studien vorgefundenen Ein- und Ausschlusskriterien tatsächlich auch in der elektronischen Routinedokumentation von fünf Universitätskliniken in Deutschland enthalten sind. Dabei zeigte sich, dass ungefähr die Hälfte der in einzelne zu dokumentierende Einheiten aufgeschlüsselten Ein- und Ausschlusskriterien auch im KIS dokumentierbar war. Allerdings waren diese Dokumentationsfelder nur bei ca. zwei Drittel der Patienten auch ausgefüllt, so dass die komplette Verfügbarkeit relevanter Daten im KIS in Bezug auf die ausgewählten Studien bei 35% lag. Während die Verfügbarkeit der Angaben zum Alter, zum Geschlecht und zur Behandlungsdiagnose wesentlich höher war, gab es z. B. deutlich weniger verfügbare Daten zu Nebenerkrankungen und der Medikation [39]. Da diese Analyse aus dem Jahr 2013 stammt, könnte sich die Vollständigkeit der Dokumentation in vielen KIS aber mittlerweile weiter entwickelt haben, so dass eine erneute Analyse zu empfehlen ist.

Auch wenn somit offensichtlich längst nicht alle für die Forschung wichtigen Daten auch schon elektronisch dokumentiert aus dem Behandlungsfall heraus vorliegen und

---

11 Siehe [www.tmf-ev.de/kisrek](http://www.tmf-ev.de/kisrek).

selbst wenn die relevanten vorhandenen Daten auch noch unpassend oder gar nicht standardisiert kodiert sind, so gibt es trotzdem doch eine Reihe von Anwendungsfällen wie z. B. auch die Rekrutierungsunterstützung, die von der Nutzung dieser Daten sehr profitieren können [39]. Es gibt bisher aber leider nur wenige Projekte, in denen die Erhöhung der Rekrutierungsrate [6] oder die Effizienzsteigerung bei der Rekrutierung bzw. Datenerhebung auch explizit nachgewiesen werden konnte [40].

Die in klinischen Studien eingesetzten Kodiersysteme werden zunehmend durch internationale Standards und hier insbesondere entsprechende Vorgaben der zuständigen Zulassungsbehörden geprägt, wobei der FDA hier eine Sonderrolle zukommt [vgl. 41]. Inwiefern diese Entwicklung auch Konsequenzen für die Standardisierung und Verwendung von Terminologien im Bereich der Dokumentation der klinischen Behandlungsdaten haben wird, ist noch nicht abschließend abschätzbar. Eine entscheidende Rolle wird dabei die Abstimmung der verschiedenen beteiligten internationalen Standardisierungsorganisationen und letztlich auch die Ausrichtung der FDA spielen. Erfreulich für die Verzahnung von Forschung und Versorgung ist der Umstand, dass die FDA in jüngerer Zeit zunehmend auf Standards auf der Versorgung verweist und in diesem Zuge z. B. die Nutzung von LOINC für die Dateneinreichung aus klinischen Studien empfiehlt [42].

Neben der Frage der geeigneten Kodierung bzw. eines notwendigen Mappings [43-45] der Daten spielen aber auch Fragen der geeigneten Datenflüsse und ihrer syntaktischen Strukturierung für die Nachnutzung eine entscheidende Rolle. Hier sind geeignete Datenstandards und Schnittstellen der beteiligten Softwaresysteme gefragt. Auf der Forschungsseite spielen in der Welt der klinischen Studien die CDISC-Standards die derzeit dominierende Rolle [46, 47], auch wenn gerade in den akademischen Forschungseinrichtungen noch eine weitergehende Standardisierung wünschenswert wäre [48]. Der vorherrschend genutzte Kommunikationsstandard im Bereich der Versorgung in Deutschland ist nach wie vor HL7 Version 2, wobei auf der Basis der Clinical Document Architecture (CDA) auch zunehmend HL7 Version 3 zum Einsatz kommt [49]. Seit 2011 hat sich zudem mit enormem Tempo das modulare, webbasierte und sehr pragmatisch ausgerichtete Konzept der Fast Healthcare Interoperability Resources (FHIR) von HL7 neben den anderen Standards etabliert.<sup>12</sup> Die bisher hier sichtbare Dynamik spricht für eine schnelle Verbreitung dieses Ansatzes und damit auch eine schnell wachsende Bedeutung für Schnittstellen zur Sekundärnutzung klinischer Daten [vgl. 50]. Zudem passt dieser Standard gut zu dem Trend der Nutzung von „leichtgewichtigen“ Clients wie etwa Apps im Gesundheitswesen und deren Anbindung über einfache Schnittstellen, aufbauend auf modernen Web-Standards.

---

<sup>12</sup> Siehe <http://hl7.org/fhir>.

Seit 2008 arbeiten die beiden Standardisierungsorganisationen HL7 und CDISC im Joint Initiative Council for Standards Harmonization (JIC)<sup>13</sup> zusammen mit der ISO und anderen Institutionen gemeinsam an der Weiterentwicklung der relevanten Standards, so dass auf Basis dieser Kooperation auch auf ein langsames Zusammenwachsen der Standards beider Welten, der Versorgung und der Forschung, gehofft werden darf. Allerdings ist erstens festzustellen, dass die CDISC-Standards jenseits regulierter klinischer Studien derzeit noch kaum zum Einsatz kommen. Gerade in der Grundlagen- und der epidemiologisch orientierten Forschung gibt es bisher keinen vergleichbar prägenden einheitlichen Standard. Zweitens ist mit einem kurzfristigen „Zusammenwachsen“ der HL7- und CDISC-Standards im Zuge der Kooperationsbemühungen auf der Ebene des JIC nicht zu rechnen und drittens gehört zu einer vollständigen Analyse der Interoperabilitäts voraussetzungen auch ein Verständnis der zu unterstützenden Prozesse.

Diese Prozesse im Sinne einer Profilierung auf der Basis vorhandener Standards und ggf. mit Hilfe zusätzlich notwendiger Rahmenvorgaben und Parametrierungen abzubilden, hat sich die Initiative Integrating the Healthcare Enterprise (IHE)<sup>14</sup> auf die Fahnen geschrieben. Aktuell gibt es entsprechende Bestrebungen, im Bereich der IHE-Domäne Quality, Research and Public Health (QRPH) und in Zusammenarbeit mit verschiedenen anwendungsnahen Projekten das Angebot an Profilen für die weitere Verzahnung von Forschung und Versorgung auszubauen [51]. In dem Zuge ist auch ein White Paper zum Secondary Data Access entstanden [52]. Auf nationaler Ebene gibt es seit 2014 eine Kooperation von IHE Deutschland mit der TMF, die ebenfalls zum Ziel hat, die Methodik und das Know-how von IHE stärker für die Forschung nutzbar zu machen.<sup>15</sup> Aktuell wird in diesem Rahmen an der konkreten Nutzung von IHE-Profilen für die Datenkommunikationsprozesse rund um Forschungsregister gearbeitet; hieraus resultierende Profilergänzungen sollen in den internationalen Standardisierungsprozess eingebracht werden.

## Kategorisierung der IT-Landschaft

Je nachdem, welchen Anwendungsfall der Sekundärnutzung klinischer Daten man sich anschaut, sind ganz unterschiedliche Systeme davon betroffen und werden andere Schnittstellen benötigt. Kurz: Man betrachtet eine jeweils andere IT-Architektur. Für viele der genannten Anwendungsfälle sind die Daten in einer mehr oder minder standardisierten Form zur Verfügung zu stellen. Dies kann mit Hilfe standardisierter Schnittstellen oder auch in speziell dafür geschaffenen Data Warehouses, welche insbesondere für Machbarkeitsstudien sehr erfolgreich eingesetzt werden, geschehen.

---

13 Joint Initiative Council: Joint Initiative on SDO Global Health Informatics Standardization, siehe [www.jointinitiativecouncil.org](http://www.jointinitiativecouncil.org).

14 Siehe [www.ihe.net](http://www.ihe.net).

15 Die Ergebnisse eines ersten gemeinsam durchgeführten Workshops finden sich unter [www.tmf-ev.de/news/1580](http://www.tmf-ev.de/news/1580).

Eine grundlegende Eigenschaft aller Architekturansätze ist somit der Datenaustausch über Anwendungsgrenzen bzw. Anwendergruppen hinweg. In einem aktuellen Entwurf eines White Papers von IHE zum Secondary Data Access wird entsprechend von „Communities“ gesprochen. Der zugehörige Architekturvorschlag sieht eine zentrale, neutrale und vertrauenswürdige Management-Instanz vor, die als Trusted Third Party (TTP) alle Anfragen entgegennimmt, prüft und an die beteiligten Communities bzw. deren Anwendungen weiterreicht. Die von diesen Anwendungen zurückgegebenen Daten werden dann von der TTP kontextbezogen de-identifiziert, ggf. aggregiert und an die anfragende Stelle weitergereicht [52]. Ein Vorteil dieses Ansatzes ist die bereits heute gegebene weitreichende Unterstützung durch vorhandene IHE-Profile. Allerdings wird auch darauf hingewiesen, dass in diesem Konzept die TTP zumindest kurzfristig Zugriff sowohl auf die medizinischen als auch auf identifizierende Daten hat, was nach anderen Pseudonymisierungskonzepten, wie sie z. B. auch von der TMF mit den deutschen Datenschutzaufsichtsbehörden konsentiert wurden [18], nicht empfohlen wird. Entsprechend wird in dem Entwurf des White Papers auch auf noch fehlende Profile eingegangen, mit deren Hilfe sich ein komplexeres ID-Management mit einer TTP, die keine medizinischen Daten sehen darf, realisieren ließe [52, S. 35ff].

Wenn man die Vogelperspektive verlässt und versucht, einen komplexen Anwendungsfall wie etwa die Rekrutierungsunterstützung umzusetzen, fällt schnell auf, dass dies eine komplexe Verzahnung von Akteuren und Prozessschritten sowohl im Versorgungs- als auch im Forschungskontext erfordert. Daher wird im Folgenden beispielhaft eine detaillierte Architektur skizziert, die entlang des Szenarios der Rekrutierungsunterstützung in klinischen Forschungsprojekten entwickelt wurde. Diese geht zurück auf ein BMBF-gefördertes Projekt zur KIS-basierten Rekrutierungsunterstützung, in dem eine ausführliche Analyse der beteiligten Systeme und notwendiger Funktionen durchgeführt wurde. Die folgenden Ausführungen basieren ganz wesentlich auf diesen Ergebnissen [2, 53, 54].

## **Studienverwaltung**

Zunächst müssen die für ein Forschungsprojekt relevanten Parameter für die Rekrutierung erhoben, gespeichert und verwaltet werden. Zu diesen von einem Studienverwaltungsmodul zu verarbeitenden Daten gehören im Wesentlichen die eindeutige Bezeichnung der Studie, die Ein- und Ausschlusskriterien sowie das zeitliche Abfrageschema und der Zeitraum der Studiendurchführung. Grundsätzlich sind diese Daten auch in einem Clinical Trial Management System (CTMS)<sup>16</sup> vorhanden, wenn dieses

---

<sup>16</sup> Siehe hierzu auch das Kapitel zu klinischen Studien im vorliegenden Report.

zum Einsatz kommt. Ob dieses auch für die hier genannten Zwecke genutzt werden kann hängt aber auch von der technischen und organisatorischen Zugänglichkeit dieser Daten ab. Somit können diese Daten ggf. auch in einem separaten Softwaremodul oder mit Hilfe von KIS-Funktionen verwaltet werden.

## **Datenbasis**

Für jeden Anwendungsfall wird eine passende Datenbasis für die Auswahl geeigneter Patienten benötigt. Als Lieferanten interessierender Daten kommen prinzipiell alle in der Versorgung eingesetzten Dokumentationssysteme in Frage. Allerdings gibt es in einem größeren Klinikum regelmäßig viele Spezialexsysteme, die zwar detaillierte Datensätze enthalten, jedoch meist nur für einen kleinen Teil der insgesamt behandelten Patienten. Daher sind für die meisten Anwendungsfälle zur Sekundärnutzung die Kernsysteme der klinischen Informationssysteme als Datenlieferanten von größerem Interesse. Diese übernehmen typischerweise zentrale abrechnungs- und dokumentationsrelevante Aufgaben und verarbeiten bzw. speichern damit Datensätze zu allen Patienten. Zu den Kernsystemen eines KIS gehören das Klinische Arbeitsplatzsystem (KAS) und die Patientendatenverwaltung (PDV). Das KAS zeichnet sich dadurch aus, dass es den Großteil der in einem Klinikum erfassten administrativen und klinischen Patientendaten verfügbar macht, während das PDV-System dazu patientenführend ist, d.h. Patienten- und Fallnummern zentral vergibt, die administrativen Patientenbewegungen verwaltet und alle abrechnungsrelevanten Daten bereit stellt. Ein weiteres Kernsystem ist der regelmäßig eingesetzte Kommunikationsserver, der alle Systeme untereinander verbindet und somit auch Daten zu allen Patienten zur Verfügung stellen kann.

Technisch lässt sich eine Datenbasis für die Sekundärnutzung sowohl als dauerhaft nutzbare und parametrierbare Abfrageschnittstelle wie auch als eigene und regelmäßig aktualisierte Datenbank<sup>17</sup> realisieren. Die vielen unterschiedlichen technischen Lösungsvarianten müssen letztlich alle eine standardisierte Abfrageschnittstelle für die Forschungsfragestellungen bieten und dürfen den Routinebetrieb des KIS nicht beeinträchtigen. Welche Variante in einer konkreten Systemumgebung zu wählen ist, hängt einerseits von den hierfür vom KIS bereitgestellten Funktionen und andererseits von dem Anwendungsfall und hier insbesondere der notwendigen zeitlichen Genauigkeit der abgefragten Informationen ab. So hilft beispielsweise die erst nach einer nächtlichen Synchronisation in einem Data Warehouse vorhandene Information über

---

17 Ein versorgungsnah und administrativ genutztes Data Warehouse, welches umfassende klinische Daten enthält, kann auch als Kernsystem angesehen werden. Bisher werden als Datenbasis für die Forschung regelmäßig aber eigenständige und für diesen Zweck spezialisierte Data-Warehouse-Systeme eingesetzt. Grundsätzlich könnten die Anforderungen aus der Versorgung und der Forschung auch von einem einzigen System bedient werden.

einen ambulant behandelten Patienten vom Vortag nicht, da dieser dann zum Zeitpunkt der möglichen Ansprache das Krankenhaus schon wieder verlassen hat. Andererseits kann ein vom Routinebetrieb weitgehend entkoppeltes Data Warehouse mit ganz eigenem Datenschema in den Anwendungsfällen die erste Wahl sein, in denen es auf eine KIS-unabhängige und standortübergreifende Standardisierung der Schnittstellen und Abfragemöglichkeiten ankommt.<sup>18</sup>

### **Abfragemodul**

Die Formulierung von Abfragen in einer technisch nutzbaren Form sowie deren Verwaltung geschieht in einem Abfragemodul. Für einmalige Abfragen z. B. im Rahmen einer Machbarkeitsanalyse erscheint ein eigenständiges Modul hierfür möglicherweise verzichtbar. Die Rekrutierungsunterstützung ist jedoch auf die wiederkehrende Ausführung der Abfragen angewiesen, so dass hierfür auch ein Zeitplan oder auslösende Events definiert und mit der Abfrage gespeichert werden müssen. Zudem sollte erkennbar sein, ob eine Abfrage schon oder noch aktiv ist, je nach Start- und Endpunkt des Rekrutierungszeitraums. Zum Testen und Anpassen von Abfragen sollte die Möglichkeit bestehen, unmittelbar Ergebnisse zu sehen, wobei diese idealerweise pseudonymisiert sein sollten, da für die Testung die Identität der betroffenen Patienten unerheblich ist. Allerdings muss zumindest für das behandelnde Personal auch die Möglichkeit bestehen, testweise in die zugehörigen Patientenakten Einblick zu nehmen, um zu verifizieren, ob die gefundenen Patienten auch tatsächlich den gewünschten Kriterien entsprechen. Wünschenswert ist zudem eine bei der Formulierung der Abfrage unterstützende Darstellung der abfragbaren Daten, z. B. in Form eines hierarchisch strukturierten Merkmalskatalogs oder auch durch eine für Ärzte verständliche „Abfragesprache“ [55]. Die Erfahrung aus bisherigen Projekten hat gezeigt, dass die Umsetzung der freitextlich formulierten Ein- und Ausschlusskriterien eines Studienprotokolls in die syntaktisch und semantisch korrekte technische Form eines Abfragemoduls eine sehr anspruchsvolle Aufgabe darstellt, die sowohl technisch-informatisches als auch medizinisches Know-how erfordert. Jede mögliche Unterstützung durch ein geeignetes Benutzerinterface ist daher von großem Wert [56].

### **Benachrichtigungsmodul**

Im Szenario der KIS-unterstützten Rekrutierung müssen die mit Hilfe einer Abfrage gefundenen Ergebnisse mit möglichen Kandidaten für ein klinisches Forschungsprojekt an die passenden Behandler übermittelt werden. Für die Benachrichtigung kommen

---

18 Vgl. hierzu auch das Kapitel zum Forschungsdatenmanagement im vorliegenden Report.

je nach Anwendungsfall und vorhandener Infrastruktur unterschiedliche technische Umsetzungsvarianten in Frage: Hier können Pager, Mobiltelefone, E-Mail oder auch Benachrichtigungsfunktionen innerhalb des KIS verwendet werden. Bei unsicheren Kommunikationskanälen ist zu beachten, dass nur eine pauschale Information (ohne Patientenbezug) über das Vorliegen potentieller Studienteilnehmer übermittelt wird und die eigentlichen Patientendaten dann im KIS abgerufen werden müssen.

## **Screeningliste**

In einer klinischen Studie muss im Regelfall eine Liste der zwecks Rekrutierung angesprochenen Patienten geführt werden, auf der auch das Rekrutierungsergebnis vermerkt ist. Zu dem Rekrutierungsergebnis gehört dann nicht nur die Information über das Vorliegen der informierten und freiwilligen Einwilligung des Patienten, sondern darüber hinaus ggf. auch das Ergebnis einer Screening-Untersuchung, die das Vorliegen der Ein- und das Nicht-Vorliegen der Ausschlusskriterien dokumentiert.

Die Screeningliste wird mit den Ergebnisdaten aus dem Abfragemodul gefüllt. Wichtig ist aber auch die Kommunikation von der Screeningliste zurück zum Abfragemodul, damit bereits einmal gefundene und ggf. angesprochene Patienten nicht erneut gefunden und wiederholt kontaktiert werden. In der Screeningliste werden die Patienten zwecks Nachvollziehbarkeit mit den vollständigen identifizierenden Angaben geführt. Damit behandelnde Ärzte eine vollständige Verifikation der Ein- und Ausschlusskriterien für die vorgeschlagenen Patienten möglichst effizient durchführen können, sollte aus der Screeningliste heraus ein direkter Aufruf der elektronischen Krankenakte eines Patienten möglich sein. Entsprechend verbleibt diese Liste in der behandelnden Einrichtung und der Zugriff darauf ist passend abzusichern.

## **Aspekte der Vernetzung**

Das Projekt EHR4CR ist ein Beispiel dafür, dass solche Infrastrukturen zur Sekundärnutzung klinischer Daten auch einrichtungsübergreifend geplant und aufgebaut werden können [5, S. 57-59]. Dies macht die vorgestellte Architektur nicht obsolet. Vielmehr kann sie helfen, schnell deutlich zu machen, an welchen Stellen standardisierte Schnittstellen zwischen den einzelnen Modulen benötigt werden. So ist offensichtlich, dass in einem übergreifenden Szenario der Rekrutierungsunterstützung die Datenbasis, das Benachrichtigungsmodul und die Screeningliste lokal betrieben werden müssen, während die Studienverwaltung und das Abfragemodul zentralisiert werden können.

## Zusammenfassung

Für die hier skizzierte Architektur mit den modular verteilten Einzelfunktionen wird man in dieser Form derzeit nur an wenigen deutschen Universitätsklinikreale Implementierungen vorfinden. Diese generische Rahmenkonzeption, kann aber helfen, vorhandene oder anzuschaffende Infrastrukturkomponenten auf notwendige Funktionen hin zu überprüfen. Innerhalb des bereits genannten Projekts zur KIS-basierten Rekrutierungsunterstützung wurden diese Funktionen an den unterschiedlichen Standorten mittels verschiedener, teilweise vorhandener, teilweise aber auch neu entwickelter Softwarekomponenten abgebildet. Als Datenbasis diene beispielsweise an einem Standort das KAS mit einem separaten Tabellenbereich und an einem anderen ein vollständig separates Forschungs-Data-Warehouse [2]. Bei einem dritten Partner übernahm gar der Kommunikationsserver die Rolle der Datenbasis [60].

Auch wenn diese modulare Architektur für die Rekrutierungsunterstützung entwickelt wurde, kann sie in Teilen auch die Unterstützung anderer Anwendungsfälle der Sekundärnutzung klinischer Daten abbilden. So wird man auch für die Machbarkeitsprüfung (Feasibility Study) eine Studienverwaltung, eine Datenbasis und ein Abfragemodul mit ähnlichen Funktionen benötigen. Lediglich die Ergebnisaufbereitung wird im Regelfall aus Gründen der Datensparsamkeit in Form anonymer Fallzahlen erfolgen, die dann auch einrichtungsübergreifend zusammengeführt werden können. Ähnliche Parallelen gelten für epidemiologische Auswertungen wie z. B. retrospektive Studien oder auch Fragestellungen aus der Versorgungsforschung.

## Aktueller Bestand

Bezüglich der IT-Ausstattung in klinischen Studien und Registern, die letztlich für die Umsetzung von Anwendungsfällen der Sekundärnutzung klinischer Daten ebenfalls relevant ist, wird hier auf die entsprechenden Kapitel in diesem Bericht verwiesen. Aber auch zum aktuellen Stand von IT-Infrastrukturen in der Versorgung kann hier kein umfassender Überblick aller forschungsrelevanten behandelnden Einrichtungen in Deutschland gegeben werden. Gerade die Infrastrukturen an den universitären Standorten mit ihren vielen Fachabteilungen und entsprechend spezialisierten Dokumentationssystemen neigen schon für sich zu enormer Komplexität [vgl. 53].<sup>19</sup> Gleichwohl sind verlässliche und aktuelle Angaben zu den eingesetzten IT-Systemen für die Planung standortübergreifender Vernetzungsmaßnahmen für die Forschung unerlässlich.

---

19 Einen Einblick in die Vielfältigkeit der IT-Systeme gibt die im Rahmen des BMBF-geförderten KIS-Rekrutierungsprojekts durchgeführte Analyse der bestehenden Infrastruktur der fünf klinischen Partner, die im Abschlussbericht zu API unter [www.tmf-ev.de/kisrek](http://www.tmf-ev.de/kisrek) dokumentiert ist.

Im Rahmen des bereits genannten und vom BMBF geförderten Projekts zur Untersuchung der KIS-basierten Rekrutierungsunterstützung wurde eine Erhebung zu den in Deutschland verbreiteten Kernsystemen an den universitären und nicht-universitären Krankenhäusern durchgeführt. Während für die Universitätskliniken eine Vollerhebung möglich war, basieren die Daten zu den nicht-universitären Krankenhäusern auf einer Stichprobenerhebung [53]. Demnach dominieren wenige Softwaresysteme für den klinischen Arbeitsplatz den deutschen Markt. Dies gilt insbesondere, wenn man sich auf die Situation in den Universitätskliniken beschränkt. Die vier Systeme i.s.h.med, ORBIS, medico und Soarian der beiden Anbieter Cerner<sup>20</sup> und Agfa decken 88% dieses Markts ab. Im KIS-Rekrutierungsprojekt waren genau diese vier Systeme und an einem fünften Standort noch eine Eigenentwicklung vertreten. Aber auch wenn man nicht-universitäre Krankenhäuser mit einbezieht, sind diese vier Systeme nach einer nicht-repräsentativen Stichprobenerhebung noch in etwas mehr als der Hälfte der Krankenhäuser vertreten.

Die weitere Analyse im Projekt ergab, dass keines der Systeme aktuell eine spezifische Unterstützung der klinischen Forschung z. B. im Sinne eines Rekrutierungsmoduls anbietet. Gleichwohl verfügen alle untersuchten Systeme über einzelne generische Funktionen, die im Sinne der weiter oben skizzierten Architektur für die Rekrutierungsunterstützung genutzt werden können. Hierzu gehörten neben der Workflow-Unterstützung insbesondere Abfrage- und Datenzugriffsmöglichkeiten sowie Berichts-Generatoren. Somit können Abfragemodule, Benachrichtigungsmodule und Screeninglisten grundsätzlich auch in bestehenden Informationssystemen der Versorgung umgesetzt werden. Allerdings werden diese generischen Funktionen teils auch schon für andere Workflows und Anwendungsfälle eingesetzt und stehen dann für die Rekrutierungsunterstützung nicht oder zumindest nicht uneingeschränkt zur Verfügung [53].

Wie bereits weiter oben ausgeführt, kann als Datenbasis für verschiedene Sekundärnutzungsszenarien auch ein separates Data Warehouse eingesetzt werden. Wenn ein solches Data Warehouse nahezu beliebige medizinische Daten aufnehmen und verwalten können soll und zudem möglichst vielen Ärzten und Forschern in einem Klinikum möglichst einfach Abfragen erlauben können soll, dann ist hierfür eine spezielle Infrastruktur notwendig, die sich typischerweise von den schon länger eingesetzten und weiter verbreiteten Data Warehouses in den administrativen Abteilungen der Krankenhäuser ganz wesentlich unterscheidet.

---

20 Im August 2014 wurde Siemens Health Services mit allen Krankenhausinformationssystemen vom amerikanischen Unternehmen Cerner übernommen.

Die TMF hat seit 2012 eine Projektserie zum Aufbau eines Integrated Data Repository Toolkit (IDRT) gefördert, die gerade akademischen Standorten die einfache Einrichtung einer Datenbasis für unterschiedliche Zielstellungen der Sekundärnutzung ermöglicht [61].<sup>21</sup> Als Basis des Werkzeugkastens wurde nach einer Evaluation im Rahmen des IT-Strategie-Projekts der TMF die Plattform i2b2<sup>22</sup> gewählt, eine vom NIH geförderte und an der Harvard University in Boston entwickelte Open-Source-Plattform, die gerade die Integration heterogener Daten mit einer einfachen benutzerfreundlichen Abfrageschnittstelle verbindet [62]. Seit 2009 hat sich die TMF auch am Aufbau einer zunächst nationalen und mittlerweile europäischen User Group der i2b2-Anwender im akademischen Umfeld beteiligt [63]. In den durch die TMF geförderten IDRT-Projekten wurde nicht nur ein Wizard für eine deutlich vereinfachte Installation eines i2b2-Systems entwickelt, sondern darüber hinaus wurden auch noch standardisierte ETL-Prozesse zur Extraktion, Aufbereitung und dem Einladen von Routinedaten, wie beispielsweise dem von allen Krankenhäusern elektronisch bereitzustellenden Datensatz gemäß § 21 des Krankenhausentgeltgesetzes (KHEntgG), bereit gestellt [64]. Nicht zuletzt haben diese Projekte auch die Etablierung der deutschen und europäischen i2b2 User Group mit unterstützt. Die i2b2 Plattform wird mittlerweile in Deutschland an ca. zehn universitären Standorten eingesetzt. Eine detaillierte Erhebung hierzu wäre wünschenswert, auch im Sinne einer Evaluation des Erfolgs der IDRT-Projekte. Um eine nachhaltige Zusammenarbeit zwischen der deutschen i2b2-Community und den Bostoner Entwicklern zu gewährleisten wurde 2014 eine Kooperationsvereinbarung zwischen der TMF und dem i2b2 Center for Biomedical Informatics in Boston geschlossen.

Wenn klinische Standorte für die Aufbereitung der Daten zur Sekundärnutzung vermehrt die Plattform i2b2 nutzen, dann könnten künftig auch standortübergreifende Auswertungen z. B. für Feasibility-Abfragen aufgrund der einheitlichen Schnittstellen zumindest für diese Standorte vereinfacht umgesetzt werden. Mit dem Shared Health Research Information Network (SHRINE)<sup>23</sup> existiert für ein solches standortübergreifendes Szenario bereits seit längerem ein technischer Lösungsansatz. Ob die Entscheidung des Projekts EHR4CR, für einen solchen übergreifenden Zugriff alternative Standards und Tools zu entwickeln [57, 58], die Forschung in Europa bestmöglich unterstützt, ist vor diesem Hintergrund kritisch zu hinterfragen. Es wird eine wichtige Aufgabe des aus diesem Projekt heraus gegründeten europäischen Non-Profit-Instituts „The European Institute for Innovation through Health Data“<sup>24</sup> sein, diese europäische Strategie mit anderen Initiativen und Standardisierungsbemühungen abzugleichen [65].

---

21 Siehe [www.tmv-ev.de/idrt](http://www.tmv-ev.de/idrt).

22 Informatics for Integrating Biology and the Bedside, siehe [www.i2b2.org](http://www.i2b2.org).

23 Siehe <https://www.i2b2.org/work/shrine.html>.

24 Siehe <http://www.i-hd.eu>.

## Bewertung und Handlungsbedarf

Zunächst ist festzustellen, dass es eine Reihe sinnvoller Anwendungsfälle der Sekundärnutzung klinischer Daten im Kontext sehr unterschiedlicher Forschungsfragestellungen gibt. Allerdings gibt es bisher für Anwendungsfälle zur Unterstützung klinischer Studien kaum verlässliche Daten zu dem durch solche Lösungen generierten ökonomischen Einsparpotential bzw. den möglichen Beschleunigungseffekten. Bisherige Analysen [z. B. 40, 66, 67] ergeben noch nicht die harte empirische Evidenz, die hier zu fordern ist.

Die Sekundärnutzung klinischer Daten kann jedoch auch die Qualität klinischer Forschung verbessern. Sie kann sogar ethisch geboten sein, wenn sie unnötige Doppeluntersuchungen vermeiden hilft oder wissenschaftlich wichtige Studien erst ermöglicht bzw. deren Planung und Durchführung deutlich beschleunigt. Als unethisch sind hingegen klinische Studien anzusehen, die begonnen werden, obwohl eine Analyse vorhandener Daten hätte zeigen können, dass man die Rekrutierungsziele in der vorgesehenen Zeit und mit den verfügbaren Mitteln nie erreichen wird. Hier werden Patienten einem Risiko ausgesetzt, obwohl der dieses Risiko rechtfertigende wissenschaftliche Fortschritt nicht erreicht werden kann. Das dieses Problem kein rein theoretisches ist, zeigen die immer wieder erschreckend hohen Zahlen der aufgrund verfehlter Rekrutierungsziele abgebrochenen Studien (Übersicht in [6]).

Eine zentrale Rahmenbedingung für die Anwendung und Umsetzung der Sekundärnutzung klinischer Daten ist zunächst ein einheitlicher und adäquater Rechtsrahmen, der allen beteiligten Akteuren bis hin zu den betroffenen Patienten eine ausreichende Rechtssicherheit verschafft. Die uneinheitliche und z. T. auch erschreckend unklare Rechtslage in Deutschland ist vor diesem Hintergrund als hoch problematisch anzusehen. Es ist heute kaum noch verständlich, warum Patientendaten in verschiedenen Bundesländern unterschiedlich geschützt werden müssen oder warum Forschern in einem Bundesland etwas erlaubt werden kann, was in einem anderen Bundesland verboten sein muss. Hier ist dringend mehr Einheitlichkeit gefordert. Die ab Mai 2018 anwendbare neue europäische Datenschutzverordnung lässt in bestimmten Bereichen weiterhin nationale Regelungen zu, so dass der Umfang der tatsächlich erreichten Harmonisierung des europäischen Rechtsrahmens erst nach Anpassung der nationalen Gesetzgebung eingeschätzt werden kann. Gerade die Heterogenität der deutschen Landeskrankengesetze wird vor diesem Hintergrund womöglich fortbestehen. Insofern ist hier weiterhin eine Harmonisierung auf nationaler Ebene notwendig.

Obwohl in Versorgung und Forschung oft gleichartige medizinische Sachverhalte erfasst werden, ist dies den dokumentierten Daten nicht immer anzusehen. Die unter-

schiedlichen Kodiersysteme und noch häufiger das Fehlen einer standardisierten Kodierung überhaupt verhindern noch zu oft eine sinnvolle Sekundärnutzung der einmal erhobenen Daten. Durch den zu geringen Standardisierungsgrad der Dokumentation werden somit auch Versorgungsforschungsprojekte beeinträchtigt oder verhindert. Auch wenn die Nutzung von Standards wie LOINC und UCUM rasch mehr verbindliche Unterstützung erfahren könnte und sollte, ist doch auch zu berücksichtigen, dass bei der Abstimmung und Einführung von Dokumentationsstandards in der Medizin immer viele beteiligte Parteien eingebunden werden müssen. Kostenträger und Leistungserbringer allein werden im Bereich der medizinischen Dokumentation nicht weit über eine rein abrechnungsbezogene Standardisierung, die ja auch bereits weitgehend existiert, hinaus kommen. Wichtig ist daher auch die Einbeziehung der wissenschaftlichen Fachgesellschaften und der Forschungseinrichtungen. Insofern ist der Aufbau einer koordinierenden nationalen Begleit- und Steuerstruktur notwendig, in der sowohl die Versorgungs- als auch die Forschungssicht repräsentiert ist [32]. Das im jüngst abgestimmten eHealth-Gesetz aufgeführte Interoperabilitätsverzeichnis kann ein hilfreicher Schritt zur besseren Verzahnung von Forschungs- und Versorgungs-IT sein, wird aber nicht allein alle hierfür nötigen Grundlagen schaffen.

Zudem sollte sowohl in der Forschung als auch der Versorgung mehr Transparenz bezüglich der Struktur der erfassten Daten bzw. der Metadaten bestehen. Ein freier Zugang zu den Metadaten im Gesundheitswesen, z. B. auch zu den medizinischen Dokumentationsformularen, ist von zentraler Bedeutung für die Sekundärnutzung klinischer Daten und würde perspektivisch die notwendige Harmonisierung unterstützen [23].

Um die Vielzahl der unterschiedlichen Dokumentationsdomänen und Anwendungsfälle in der Medizin „unter einen Hut zu bringen“ bzw. die verschiedenen Daten aufeinander abbilden oder zueinander in Beziehung setzen zu können, ist perspektivisch der Einsatz einer Referenzterminologie unverzichtbar. Hier erscheint derzeit SNOMED-CT aufgrund der umfangreich durchgeführten Vorarbeiten und auch aktueller internationaler Aktivitäten als vielversprechendste Lösung [28]. Hierbei müssten nicht nur lizenzrechtliche und technologische Aspekte berücksichtigt werden. Vielmehr wäre die Einführung einer Referenzterminologie ein höchst komplexes Vorhaben, welches mit ausreichenden Ressourcen sowohl für IT-Systeme und deren Customizing, als auch für Anwenderschulungen und Akzeptanzmaßnahmen unterstützt werden müsste. Parallel zum schrittweisen Einstieg in die Nutzung einer Referenzterminologie ist zudem die Förderung der Forschung und des Nachwuchses auf dem Feld der medizinischen Terminologie auszubauen [32].

In der Verbundforschung besteht die Herausforderung, eine möglichst einheitliche und zentral gesteuerte IT-Infrastruktur mit den heterogenen IT-Infrastrukturen in den beteiligten klinischen Einrichtungen zu verbinden. Andersherum besteht für die lokalen Krankenhäuser das Problem, Schnittstellen zu den heterogenen Vorgaben der verschiedenen Verbundforschungsvorhaben, an denen das Haus beteiligt ist, einzurichten und zu betreuen. Beide Seiten können letztlich nur von einer zunehmenden Nutzung breit abgestimmter Standards für Schnittstellen und Datenformate profitieren. Die Möglichkeit, über die TMF einen solchen Prozess zu moderieren und zu unterstützen, insbesondere gemeinsam die Konsensbildung zur Definition und verbindlichen Festsetzung von Standards zu betreiben, bietet erkennbare Vorteile.

Die Bestandsanalyse im Bereich der Rekrutierungsunterstützung hat deutlich gemacht, dass die bestehenden IT-Systeme in der Versorgung noch zu wenig auf die parallele Unterstützung von Forschungsprozessen ausgerichtet sind. Dabei müssen nicht alle notwendigen Funktionen immer innerhalb eines Softwaresystems angeboten werden. Auch modulare Lösungen haben ihre Vorteile. Allerdings zeigt der Prozess der Rekrutierungsunterstützung gerade die Notwendigkeit einer engen Verzahnung der einzelnen benötigten Funktionen auf. Insofern muss auch bei den Herstellern der IT-Lösungen in der Versorgung mehr für die Forschungsunterstützung geworben werden. Gleichzeitig sollten forschungsunterstützende Funktionen von den klinischen Einrichtungen vermehrt angefragt werden.

Viele der genannten Aspekte und Handlungsfelder werden von dem aktuellen Förderkonzept Medizininformatik des BMBF adressiert.<sup>25</sup> Explizites Ziel dieser Initiative ist es, Austausch und Nutzung von Daten nicht nur zwischen Institutionen, sondern insbesondere auch zwischen der Krankenversorgung und der klinischen bzw. biomedizinischen Forschung zu intensivieren. Hierfür ist der Aufbau einrichtungsübergreifend anschlussfähiger Datenintegrationszentren vorgesehen, die sowohl Versorgungs- als auch Forschungsdaten in standardisierter Form zugänglich machen sollen. Die hierfür nötigen technischen, organisatorischen und rechtlichen Rahmenbedingungen werden in einer nationalen Begleitstruktur erarbeitet und abgestimmt.<sup>26</sup> Es ist zu hoffen, dass diese wichtige, die Universitätsmedizin nahezu aller Bundesländer betreffende Förderinitiative dazu genutzt werden kann, zu vielen der aufgezeigten Handlungsfelder in den nächsten Jahren deutliche Fortschritte zu verzeichnen.

---

25 Siehe <http://www.gesundheitsforschung-bmbf.de/de/6513.php>.

26 Siehe <http://www.tmf-ev.de/news/3020>.

## Literatur

1. Christoph, J., Griebel, L., Leb, I. et al., *Secure Secondary Use of Clinical Data with Cloud-based NLP Services Towards a Highly Scalable Research Infrastructure*. Methods of Informatics in Medicine, 2015. **54**(3 2015): S. 276-282.
2. Trinczek, B., Köpcke, F., Leusch, T. et al., *Design and multicentric Implementation of a generic Software Architecture for Patient Recruitment Systems re-using existing HIS tools and Routine Patient Data*. Applied Clinical Informatics, 2014. **5**(1): S. 264 - 283.
3. De Moor, G., Sundgren, M., Kalra, D. et al., *Using electronic health records for clinical research: the case of the EHR4CR project*. J Biomed Inform, 2015. **53**: S. 162-173.
4. Prokosch, H.U., Ganslandt, T., *Perspectives for medical informatics. Reusing the electronic medical record for clinical research*. Methods of Information in Medicine, 2009. **48**(1): S. 38-44.
5. Coorevits, P., Sundgren, M., Klein, G.O. et al., *Electronic health records: new opportunities for clinical research*. J Intern Med, 2013. **274**(6): S. 547-560.
6. Dugas, M., Lange, M., Müller-Tidow, C., Kirchhof, P., Prokosch, H.-U., *Routine data from hospital information systems can support patient recruitment for clinical studies*. Clinical Trials, 2010. **7**(2): S. 183-189.
7. Prokosch, H.-U., *Single-Source-Aktivitäten in Deutschland*. mdi, 2010. **12**(2): S. 56-57.
8. Swart, E., Ihle, P., Gothe, H., Matusiewicz, D., Hrsg. *Routinedaten im Gesundheitswesen*. 2 Aufl. 2014, Huber, Bern.
9. Declerck, G., Hussain, S., Daniel, C., Yuksel, M., Laleci, G.B., Twagirumukiza, M., Jaulent, M.C., *Bridging Data Models and Terminologies to Support Adverse Drug Event Reporting Using EHR Data*. Methods of Information in Medicine, 2015. **54**(1): S. 24-31.
10. Kush, R., Alschuler, L., Ruggeri, R. et al., *Implementing Single Source: the STARBRITE proof-of-concept study*. J Am Med Inform Assoc, 2007. **14**(5): S. 662-673.
11. El Fadly, A., Lucas, N., Rance, B., Verplancke, P., Lastic, P.Y., Daniel, C., *The REUSE project: EHR as single datasource for biomedical research*. Stud Health Technol Inform, 2010. **160**(Pt 2): S. 1324-1328.
12. El Fadly, A., Rance, B., Lucas, N. et al., *Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform*. J Biomed Inform, 2011. **44 Suppl 1**: S. S94-102.

13. Bruland, P., McGilchrist, M., Zapletal, E. et al., *Common data elements for secondary use of electronic health record data for clinical trial execution and serious adverse event reporting*. BMC Med Res Methodol, 2016. **16**(1): S. 159.
14. Prokosch, H.U., Beck, A., Ganslandt, T. et al., *IT Infrastructure Components for Biobanking*. Applied Clinical Informatics, 2010. **1**(4): S. 419-429.
15. Prokosch, H.-U., Mate, S., Christoph, J. et al., *Designing and implementing a biobanking IT framework for multiple research scenarios*. Stud Health Technol Inform, 2012. **180**: S. 559-563.
16. Schneider, U.K., *Sekundärnutzung klinischer Daten - Rechtliche Rahmenbedingungen*. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
17. AKmedEK Mustertext zur Spende, Einlagerung und Nutzung von Biomaterialien sowie zur Erhebung, Verarbeitung und Nutzung von Daten in Biobanken. Empfohlen vom Arbeitskreis Medizinischer Ethik-Kommissionen gemäß Beschluss vom 10.6.2016. 2016. Arbeitskreis Medizinischer Ethik-Kommissionen, Version 2.0, <http://www.ak-med-ethik-komm.de/documents/MustertextBiobanken.docx> (Abruf: 2016-12-16).
18. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
19. EuGH Urteil des Gerichtshofs (Zweite Kammer) vom 19. Oktober 2016. Patrick Breyer gegen Bundesrepublik Deutschland. Aktenzeichen C-582/14. 2016. Europäischer Gerichtshof.
20. Sariyar, M., Schlünder, I., *Reconsidering Anonymization-Related Concepts and the Term „Identification“ Against the Backdrop of the European Legal Framework*. Biopreserv Biobank, 2016.
21. Weichert, T., *Die Europäische Datenschutz-Grundverordnung – ein Überblick*. DANA, 2016. **2016**(2): S. 48-56.
22. Dugas, M., Jöckel, K.-H., Gefeller, O. et al., *Freier Zugang zu Dokumentationsformularen und Merkmalskatalogen im Gesundheitswesen. Memorandum „Open Metadata“*. GMS Medizinische Informatik, Biometrie und Epidemiologie, 2014. **10**(1).
23. Dugas, M., Jöckel, K.H., Friede, T. et al., *Memorandum „Open Metadata“*. Open Access to Documentation Forms and Item Catalogs in Healthcare. Methods of Information in Medicine, 2015. **54**(4): S. 376-378.
24. Dugas, M., Neuhaus, P., Meidt, A., Doods, J., Storck, M., Bruland, P., Varghese, J., *Portal of medical data models: information infrastructure for medical research and healthcare*. Database, 2016. **2016**.

25. Horenkamp-Sonntag, D., Lindner, R., Wenzel, F., Gerste, B., Ihle, P., *Prüfung der Datenqualität und Validität von GKV-Routinedaten*, in *Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven*, Hrsg.: E. Swart, et al. 2014, Huber, Bern. S. 314-330.
26. Rienhoff, O., *Die zunehmende Bedeutung von Terminologiesystemen in Forschung und Versorgung in einer individualisierten Medizin*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 29-34.
27. Ingenerf, J., Pöpl, S.J., *Biomedical vocabularies – the demand for differentiation*, in *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems*, Hrsg.: K.A. Kuhn, J.R. Warren, and T.-Y. Leong. 2007, IOS Press, Amsterdam. S. 610-615.
28. Ingenerf, J., Schopen, M. *Positionspapier zur „Systematized Nomenclature of Medicine - Clinical Terms“ (SNOMED CT) in Deutschland*. 2006. GMDS-Arbeitsgruppe Standardisierte Terminologien in der Medizin, <http://www.gmds.de/pdf/publikationen/stellungnahmen/Positionspapier.pdf> (Abruf: 2016-02-08).
29. Dugas, M., *Missing semantic annotation in databases. The root cause for data integration and migration problems in information systems*. *Methods of Informatics in Medicine*, 2014. **53**(6): S. 516-517.
30. Dewenter, H., Thun, S., *Die SNOMED CT-Herausforderung*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 89-96.
31. Semler, S.C., Weil, P., Bockhorst, K., Bauer, C.R.K.D., Gusky, L., Rienhoff, O., *Nutzung von Terminologien und Ordnungssystemen in der Medizin in Deutschland, Österreich und Schweiz – Stand und Perspektiven*, in *Big Data und Forschungsinfrastruktur - Perspektiven für die Medizin. Abstractband GMDS 2014*, Hrsg.: S.Y. Nußbeck, S. Wolff, and O. Rienhoff. 2014, Schattauer, Stuttgart. S. 581-583.
32. Weil, P., Gusky, L., Bauer, C., Bockhorst, K., Semler, S.C., Rienhoff, O., *Thesenpapier des D-A-CH-Projekts. Empfehlungen an das Bundesministerium für Gesundheit*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 7-20.
33. Kim, S.-I., *Nutzungsmöglichkeiten von SNOMED CT in der Schweiz*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 193-195.

34. Geßner, C., Wehnert, J., *Der epSOS Master Value Catalogue (MVC): Erfahrungen bei der Auswahl von Codesystemen*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 181-187.
35. Semler, S.C., Röhrig, R., *LOINC – Internationale Nomenklatur zur Kodierung von medizinischen Untersuchungen und Befunden*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 97-134.
36. Zunner, C., Bürkle, T., Prokosch, H.U., Ganslandt, T., *Mapping local laboratory interface terms to LOINC at a German university hospital using RELMA V.5: a semi-automated approach*. *J Am Med Inform Assoc*, 2013. **20**(2): S. 293-297.
37. Geßner, C., *Kodierung von Einheiten physikalischer Messgrößen mit UCUM*. *HL7-Mitteilungen*, 2006(21): S. 6-17.
38. Ingenerf, J., Kock, A.K., Poelker, M., Seidl, K., Zeplin, G., Mersmann, S., Handels, H., *Standardizing intensive care device data to enable secondary usages*. *Studies in Health technology and Informatics*, 2012. **180**: S. 619-623.
39. Köpcke, F., Trinczek, B., Majeed, R.W. et al., *Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence*. *BMC Med Inform Decis Mak*, 2013. **13**: S. 37.
40. Köpcke, F., Kraus, S., Scholler, A., Nau, C., Schuttler, J., Prokosch, H.U., Ganslandt, T., *Secondary use of routinely collected patient data in a clinical trial: an evaluation of the effects on patient recruitment and data acquisition*. *Int J Med Inform*, 2013. **82**(3): S. 185-192.
41. *FDA Guidance for Industry. Providing Regulatory Submissions In Electronic Format - Standardized Study Data*. 2014. U.S. Department of Health and Human Services. Food and Drug Administration, <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM292334.pdf> (Abruf: 2016-01-18).
42. *FDA, Electronic Study Data Submission; Data Standards; Support for the Logical Observation Identifiers Names and Codes*. Federal Register, 2015. **80**(93): S. 27690-27691 (e-Pub: 05/14/2015), <https://www.gpo.gov/fdsys/pkg/FR-2015-05-14/pdf/2015-11596.pdf> (Abruf: 2016-12-04).
43. Ouagne, D., Hussain, S., Sadou, E., Jalent, M.C., Daniel, C., *The Electronic Healthcare Record for Clinical Research (EHR4CR) information model and terminology*. *Stud Health Technol Inform*, 2012. **180**: S. 534-538.
44. Hussain, S., Sun, H., Sinaci, A. et al., *A framework for evaluating and utilizing medical terminology mappings*. *Stud Health Technol Inform*, 2014. **205**: S. 594-598.

45. Mate, S., Köpcke, F., Toddenroth, D., Martin, M., Prokosch, H.U., Bürkle, T., Ganslandt, T., *Ontology-based data integration between clinical and research systems*. PLoS One, 2015. **10**(1): S. e0116656.
46. de Montjoie, A.J., *CDISC-Standards: reaching out to a global audience*. World Pharma Network, 2008: S. 10-12.
47. Richesson, R.L., Nadkarni, P., *Data standards for clinical research data collection forms: current status and challenges*. J Am Med Inform Assoc, 2011. **18**(3): S. 341-346.
48. Kuchinke, W., Ohmann, C., Yang, Q. et al., *Heterogeneity prevails: the state of clinical trial data management in Europe - results of a survey of ECRIN centres*. Trials, 2010. **11**: S. 79-89.
49. Dolin, R.H., Rogers, B., Jaffe, C., *Health Level Seven Interoperability Strategy: Big Data, Incrementally Structured*. Methods of Information in Medicine, 2015. **54**(1): S. 75-82.
50. Doods, J., Neuhaus, P., Dugas, M., *Converting ODM Metadata to FHIR Questionnaire Resources*. Stud Health Technol Inform, 2016. **228**: S. 456-460.
51. Daniel, C., Erturkmen, G.B., Sinaci, A.A., Delaney, B.C., Curcin, V., Bain, L., *Standard-based integration profiles for clinical research and patient safety*. AMIA Jt Summits Transl Sci Proc, 2013. **2013**: S. 47-49.
52. IHE *Using IHE profiles for Healthcare - Secondary Data Access. IHE Quality, Research and Public Health (QRPH) White Paper*. 2016. Integrating the Healthcare Enterprise, Draft for Public Comment. Revision 1.0, [https://www.ihe.net/uploadedFiles/Documents/QRPH/IHE\\_QRPH\\_WP\\_Healthcare\\_Secondary\\_Data\\_Access\\_Rev1.0\\_PC\\_2016-07-25.pdf](https://www.ihe.net/uploadedFiles/Documents/QRPH/IHE_QRPH_WP_Healthcare_Secondary_Data_Access_Rev1.0_PC_2016-07-25.pdf) (Abruf: 2016-11-27).
53. Schreiweis, B., Trinczek, B., Köpcke, F. et al., *Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials*. Int J Med Inform, 2014. **83**(11): S. 860-868.
54. Trinczek, B., Schulte, B., Breil, B., Dugas, M., *Patient recruitment workflow with and without a patient recruitment system*. Stud Health Technol Inform, 2013. **192**: S. 1124.
55. Bache, R., Taweel, A., Miles, S., Delaney, B.C., *An Eligibility Criteria Query Language for Heterogeneous Data Warehouses*. Methods of Information in Medicine, 2015. **54**(1): S. 41-44.
56. Doods, J., Holzapfel, K., Dugas, M., Fritz, F., *Development of best practice principles for simplifying eligibility criteria*. Stud Health Technol Inform, 2013. **192**: S. 1153.
57. Doods, J., Bache, R., McGilchrist, M., Daniel, C., Dugas, M., Fritz, F., *Piloting the EHR4CR Feasibility Platform across Europe*. Methods of Information in Medicine, 2014. **53**(4): S. 264-268.

58. Soto-Rey, I., N'Dja, A., Cunningham, J. et al., *User Satisfaction Evaluation of the EHR4CR Query Builder: A Multisite Patient Count Cohort System*. Biomed Res Int, 2015. **2015**: S. 801436.
59. McCowan, C., Thomson, E., Szmigielski, C.A. et al., *Using Electronic Health Records to Support Clinical Trials: A Report on Stakeholder Engagement for EHR4CR*. Biomed Res Int, 2015. **2015**: S. 707891.
60. Majeed, R.W., Röhrig, R., *Identifying patients for clinical trials using fuzzy ternary logic expressions on HL7 messages*. Stud Health Technol Inform, 2011. **169**: S. 170-174.
61. Bauer, C.R.K.D., Ganslandt, T., Baum, B. et al., *Integrated Data Repository Toolkit (IDRT). A Suite of Programs to Facilitate Health Analytics on Heterogeneous Medical Data*. Methods Inf Med, 2016. **55**(2): S. 125-135.
62. Kohane, I.S., Churchill, S.E., Murphy, S.N., *A translational engine at the national scale: informatics for integrating biology and the bedside*. J Am Med Inform Assoc, 2012. **19**(2): S. 181-185.
63. Ganslandt, T., Mate, S., Helbing, K., Sax, U., Prokosch, H.U., *Unlocking data for clinical research - The German i2b2 experience*. Applied Clinical Informatics, 2010. **1**(4): S. 116-127.
64. Ganslandt, T., Sax, U., Löbe, M. et al., *Integrated Data Repository Toolkit: Werkzeuge zur Nachnutzung medizinischer Daten für die Forschung*, in *Informatik 2012*, Hrsg.: U. Goltz, et al. 2012, Köllen Druck+Verlag GmbH, Bonn. S. 1252-1259.
65. Kalra, D., Stroetmann, V., Sundgren, M. et al., *The European Institute for Innovation through Health Data*. Learn Health Sys, 2016. **2016**: S. 1-8.
66. Beresniak, A., Schmidt, A., Proeve, J. et al., *Cost-benefit assessment of using electronic health records data for clinical research versus current practices: Contribution of the Electronic Health Records for Clinical Research (EHR4CR) European Project*. Contemp Clin Trials, 2016. **46**: S. 85-91.
67. Soto-Rey, I., Trinczek, B., Girardeau, Y. et al., *Efficiency and effectiveness evaluation of an automated multi-country patient count cohort system*. BMC Med Res Methodol, 2015. **15**: S. 44.



# Forschungsdatenmanagement

*Ulrich Sax, Christian Bauer, Thomas Ganslandt, Toralf Kirsten*

## Anwendungsbereich

Die Vokabel Forschungsdatenmanagement hat aufgrund ihrer Bedeutung mittlerweile Eingang in die Förderkonzepte nationaler und internationaler Drittmittelgeber gefunden [1, 2].<sup>1</sup> Hintergrund ist die Erkenntnis, dass viele bislang als qualitativ hochwertig geglaubte wissenschaftliche Arbeiten bei genauerer Untersuchung nicht nachvollzogen werden können [3]. Ein belastbares – jedoch durchaus mit Aufwand verbundenes – Forschungsdatenmanagement hat zur Aufgabe, jegliche Datenquelle mit klarer Herkunft, Zeitstempel, Autorenschaft etc. dauerhaft zu belegen. Die aktuellen Stellungnahmen bzw. Empfehlungen des Rats für Informationsinfrastruktur (RFII) [4], der Deutschen Forschungsgemeinschaft (DFG) [5] sowie der Hochschulrektorenkonferenz (HRK) [6] belegen die Dringlichkeit des Themas und den hier bestehenden Handlungsbedarf.

Der Begriff Forschungsdatenmanagement wird sehr vielschichtig verwendet und bezieht sich üblicherweise neben den technischen Grundlagen des Datenmanagements auch auf weitere Aspekte wie etwa Provenance (Wo kommen die Daten her, wie wurden sie verändert?), Governance (Wer beaufsichtigt die Speicherung und Nutzung der Daten?), Use and Access (Wie wird die Nutzung beantragt?), Data Transfer Unit (Wie werden die Daten bei genehmigter Nutzung selektiert, exportiert und weitergeleitet?) und Langzeitarchivierung. Dieses Kapitel behandelt vorrangig die technischen Aspekte des Managements von Forschungsdaten (Erhebung, Herauslösen aus dem Primärkontext, Zusammenführung etc.). Auf verschiedene Stellungnahmen aus dem wissenschaftsnahen Umfeld zu weiteren Aspekten des Forschungsdatenmanagements wird in dem Kapitel zu strategischen Aktivitäten und Fördermaßnahmen näher eingegangen.

Das Open Archival Information System<sup>2</sup> (ISO 14721) gilt allgemein hin als Referenzmodell für ein Forschungsdatenmanagement. Es zeigt sowohl die benötigten Komponenten und Prozesse als auch die Akteure und deren Aufgaben auf, die an einem einheitlichen Forschungsdatenmanagement teilhaben. Dazu gehören die Datenproduzenten, die Daten in verschiedenen Kontexten generieren und für ein übergreifendes Forschungsdatenmanagement zur Verfügung stellen (vgl. Datenquellen in Abb. 10.1). Dem entgegengesetzt nutzen die Datenkonsumenten den verwalteten Datenbestand,

1 Siehe auch ELIXIR: A distributed infrastructure for life-science information. <https://www.elixir-europe.org/about/eu-projects>.

2 Siehe <https://public.ccsds.org/pubs/650x0m2.pdf>.

z. B. zum Zwecke der Visualisierung, Datenexploration oder der gezielten Auswertung anhand von dedizierten medizinisch-biologischen Hypothesen. Dazu bedarf es verschiedener Komponenten, um einen Überblick über die akquirierten Daten und Projekte (z. B. Register) bis hin zu den detaillierten Metadaten einzelner Datenelemente in einem Metadaten Repository gewinnen zu können. Insofern versteht sich dieses Kapitel als Querschnittskapitel, das den aktuellen Stand und den Handlungsbedarf im Umfeld eines IT-gestützten Forschungsdatenmanagements wiedergibt.

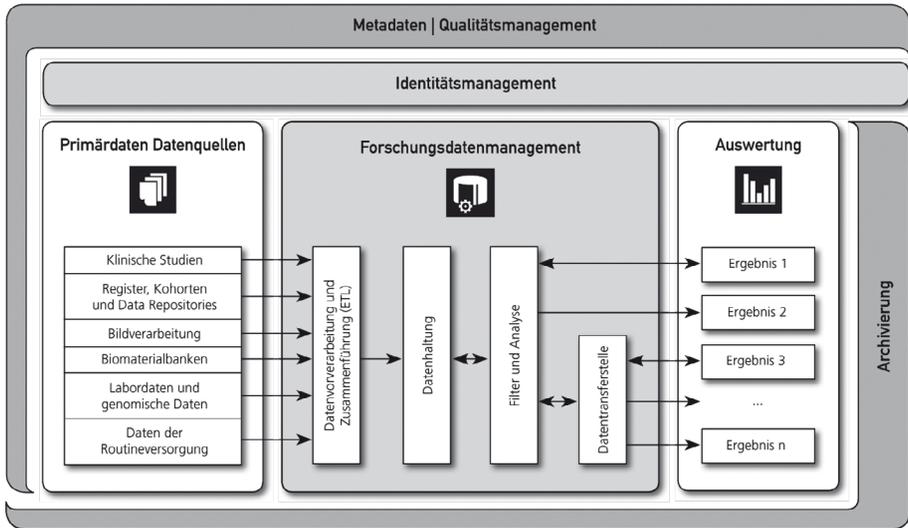


Abbildung 10.1: Generische Darstellung der Einbettung des Forschungsdatenmanagements in die übergreifende Datenverarbeitung in der patientenorientierten Forschung von den Datenquellen bis zur Auswertung und Archivierung.

Auf Seiten der Datenquellen wurden in mehreren deutschen Krankenhäusern Ansätze zur Wiederverwendung der in den elektronischen Krankenakten gespeicherten Daten für Forschungszwecke entwickelt und erfolgreich angewendet [7-9].<sup>3</sup> Diese phänotypischen Daten werden zunehmend aus der Grundlagenforschung mit Hochdurchsatzanalysen (Genomics, Proteomics, Transcriptomics, Metabolomics) von Patientenproben ergänzt. Nur die Verknüpfung der umfangreichen Ergebnisse solcher Omics-Analysen mit den klinischen Phänotypdaten erlaubt aber, das volle wissenschaftliche Potential dieser Daten zu heben.

Genau diese Integration stellt jedoch nach wie vor eine große Herausforderung in der biomedizinischen Informatik dar. Auch wenn bereits erste Werkzeuge bzw. generische

<sup>3</sup> Vgl. hierzu auch das Kapitel zur Sekundärnutzung klinischer Daten im vorliegenden Report.

Integrationsplattformen verfügbar sind, so zeigt doch ein erster Vergleich, dass noch Raum für Weiterentwicklungen besteht, insbesondere hinsichtlich der Anforderungen des Datenschutzes und der Interoperabilität [10]. Sollen Daten standortübergreifend ausgewertet werden, sollte über eine reine Extraktion und Zusammenführung von Rohdaten auch eine Vereinheitlichung erfolgen, z. B. auf Basis eines übergreifenden Datenmodells (Common Data Model).

## Kategorisierung der IT-Landschaft

### **Werkzeuge zur Extraktion, Transformation und zum Laden (ETL) von Daten**

Mit Hilfe von Extraktions-, Transformations- und Ladeprozessen (ETL) können Daten aus Quellsystemen in Ablage- und Auswertungsplattformen übertragen werden. Der Begriff entstand im Kontext von Data-Warehouse-Systemen; im Bereich des Forschungsdatenmanagements wird auch von „ingest“ oder „harvesting“ gesprochen. Der Extraktionsschritt besteht zunächst in der Erstellung einer Kopie der Daten des Quellsystems in einem sogenannten „Staging“-Bereich. Ziel dieses Schritts ist die Entkopplung der weiteren Verarbeitung von den Quellsystemen. Im Rahmen des Transformationsschritts können Formatkonvertierungen sowie Schritte zur Integration in ein gemeinsames Datenmodell erfolgen. Im Rahmen des Forschungsdatenmanagements muss hier zwischen den Zielen einer Erhaltung der Quelldaten in unveränderter Form und der optimierten Verwertung in einem übergreifenden Datenmodell abgewogen werden. Der „Loading“-Schritt beschreibt das abschließende Laden der extrahierten und ggf. transformierten Daten in eine Speicherplattform (Forschungsdaten-Repository), wobei hier Optimierungsmöglichkeiten z. B. im Sinne einer Indizierung oder Partitionierung angewendet werden können. ETL-Werkzeuge stehen sowohl von kommerziellen Anbietern als auch im Open Source-Bereich zur Verfügung. Sie stellen typischerweise eine Vielzahl von Schnittstellen zur Verfügung, mit denen Datenbanken, Dateiformate oder Services angebunden werden können. Daneben bieten Sie Möglichkeiten zum Mapping von Datenschemata sowie zur Automatisierung von Verarbeitungsschritten.

### **Datenhaltungssysteme / Forschungsdaten-Repositories**

Datenhaltungssysteme nehmen die im Rahmen von ETL-Prozessen erschlossenen Daten auf, um sie entweder für die Verwertung mit Filter- und Analysewerkzeugen bereitzustellen oder langfristig zur Rückverfolgung wissenschaftlicher Ergebnisse oder für andere zukünftige Nutzungen aufzubewahren. Typischerweise kommen relationale Datenbankmanagementsysteme (RDBMS) zum Einsatz, um strukturierte Daten zu speichern. Dazu zählen in der Regel Phänotypdaten, die in klinischen und epidemiolo-

gischen Studien oder in der Versorgungspraxis gewonnen werden. Omics-Analyseergebnisse sowie Bilddaten und andere unstrukturierte Daten können ebenso in RDBMS gespeichert werden, sind jedoch nicht oder nur eingeschränkt abfragbar. Daher verbleiben sie oftmals in Dateiform in Verzeichnissen, sind aber mit den strukturierten Daten im Datenbanksystem verknüpft. Als Alternative für variabel strukturierte – aber auch Omics-Daten – kommen zunehmend NoSQL-Datenbanken in Betracht. Ergänzend können aus diesen unstrukturierten Daten durch entsprechende Verfahren wie z. B. Text-Mining auch strukturierte Informationen gewonnen und für weitere Auswertungen mit vorgehalten werden. Für die Langzeitarchivierung stehen Aspekte der Revisionsicherheit und Berechtigungsstruktur einerseits sowie einer Verschlagwortung auf Basis von standardisierten Terminologien andererseits im Vordergrund.

### **Common Data Models (CDM)**

Insbesondere bei einer standortübergreifenden Nutzung klinischer Routinedaten ist über die reine Erschließung und Integration hinaus eine weitergehende Harmonisierung erforderlich. Diese umfasst sowohl die Vereinbarung eines einheitlichen Datenmodells als auch die Festlegung übergreifend genutzter Terminologien. Auf dieser Basis können Abfragen zentral definiert und in identischer Form standortübergreifend auf den jeweiligen lokalen Datenbeständen ausgeführt werden. Es ist zu beachten, dass es sowohl durch die Transformation in ein gemeinsames Datenmodell als auch durch die Abbildung lokaler Konzepte auf die ausgewählten Terminologien zu Informationsverlusten kommen kann. Ein CDM kann dem zumindest teilweise Rechnung tragen, indem es neben den zentral festgelegten Terminologien auch die Ablage der lokal verwendeten Originalkonzepte unterstützt. Die Dokumentation der Transformationsschritte vom lokalen Datenmodell in das CDM kann helfen, Veränderungen und mögliche Informationsverluste nachzuvollziehen. Eine strukturierte Ablage solcher Provenance-Informationen im CDM könnte darüber hinaus die Interpretation der Daten unterstützen.

### **Filter- und Analysewerkzeuge**

Wesentliche Voraussetzung für die nachhaltige Verwertung der erschlossenen Daten sind Werkzeuge, die den wissenschaftlichen Anwendern eine eigenständige Analyse ermöglichen. Sie erlauben die flexible Kombination von Ein- und Ausschlusskriterien, um Kohorten definieren zu können. Optional können auch darüber hinausgehende Auswertungs- und Visualisierungsfunktionen integriert oder Wissensbanken in die Analyse einbezogen werden. Ein generisches Datenmodell erlaubt die Nutzung mit Daten aus verschiedenen Quellen, wobei molekularbiologische Daten aufgrund ihres Volumens häufig separate Konstrukte zur Ablage und Analyse benötigen.

## Data Transfer Units

Die Auswertung der in den Studien gesammelten Daten setzt den Zugriff auf die Daten in der Studiendatenbank bzw. einem übergreifenden Forschungsdaten-Repository voraus. Der Datenzugriff ist in der Regel nicht öffentlich, sondern – um dem Datenschutz und den „intellectual properties“ der initialen Akteure Rechnung zu tragen – oftmals auf einen bestimmten Personenkreis und eine Auswahl von Daten begrenzt, bzw. an eine abgestimmte Policy gekoppelt. In größeren Studien, mit vielen und unterschiedlichen Akteuren mit je eigenen Interessen, wird die Schnittstelle zwischen Datenmanagement und Datenanalyse von einer eigenen Organisationseinheit, einer „data transfer unit“ (Datentransferstelle, DTU) ausgefüllt. Diese stellt die relevanten Daten auf Anforderung aus dem Forschungsdaten-Repository zusammen und macht sie den relevanten Personenkreisen verfügbar. Die Anforderungen sind üblicherweise formalisiert; d. h. Daten werden auf Grund eines Antrags (-Dokuments) bereitgestellt. Mit zunehmender Komplexität der Studie hilft die Formalisierung der Anforderungen bei der Durchführung und Dokumentation. In großen Kohortenstudien „NAKO Gesundheitsstudie“<sup>4</sup>, SHIP<sup>5</sup>, LIFE<sup>6</sup> u. a. findet die Formalisierung in sogenannten Datennutzungsanträgen oder Projektvereinbarungen ihren Niederschlag, über die ein wissenschaftliches Gremium berät und entscheidet. Die positive Zustimmung soll die Originalität der Fragestellung und Qualität der Datenanalyse wahren sowie der Überwachung des Datenschutzes und der Wahrung der „intellectual properties“ dienen. Die Anforderungen an ein solches Verfahren und das entsprechende Gremium aus Sicht des Datenschutzes sind in den generischen und mit allen relevanten Datenschutzbehörden in Deutschland abgestimmten Datenschutzkonzepten der TMF beschrieben [11].

Die ausschließliche Ausrichtung einer DTU auf die Datenbereitstellung greift jedoch zu kurz. Oftmals verlangt die verfolgte medizinische Fragestellung in den Auswertungsprojekten eine spezifische Analyse von Biomaterial, deren Ergebnisse mit den bereitgestellten Phänotypdaten zusammenhängend analysiert werden müssen. Aufgabe der DTU ist dabei nicht nur die Koordinierung des Probenverkehrs für die sich anschließende Messung im Labor sondern auch die Reintegration der erzeugten Labordaten in das Forschungsdaten-Repository. Dies gilt ebenso für relevante Ergebnisse, die sich im Zuge der Datenanalysen ergeben und von den betreffenden Forschungsgruppen zur Aufnahme in das Forschungsdaten-Repository zur Verfügung gestellt werden.

---

4 Siehe <http://nako.de>.

5 Siehe [https://www.fvcm.med.uni-greifswald.de/dd\\_service/data\\_use\\_intro.php](https://www.fvcm.med.uni-greifswald.de/dd_service/data_use_intro.php).

6 Siehe <http://life.uni-leipzig.de>.

Aus dieser Aufgabenskizze und den davon ableitbaren Anforderungen wird deutlich, dass DTUs in technischer Hinsicht von der Qualität der Datenhaltungssysteme sowie der Filter- und Analysewerkzeuge, z. B. zur Fallzahlermittlung, abhängig sind, um ihre Schnittstellenfunktion optimal ausfüllen zu können.

## Aktueller Bestand und Erhebungsergebnisse

Aufgrund einer Vielzahl von Open-Source-Software-Stacks nehmen wir hier eine feinere Unterteilung der Werkzeugklassen entlang der typischen Forschungsdatenpipelines vor. Die Listen zum aktuellen Bestand sind nach bestem Wissen und Gewissen erstellt. Sie sind mit hoher Wahrscheinlichkeit nicht vollständig, die Darstellung beruht aber in vielen Fällen auf persönlichen Erfahrungen bzw. Empfehlungen.

### ETL-Werkzeuge

Neben einer Reihe kommerzieller Werkzeuge zum Extrahieren, Transformieren und Laden von Daten haben sich im Umfeld der biomedizinischen Forschung einige Open-Source-Werkzeuge etabliert. Allen Werkzeugen ist gemeinsam, dass bereits viele Adaptoren für verschiedene Datenquellen verfügbar sind. Zumeist lassen sich die Extraktions- und Transformationsschritte mit einer Skriptsprache automatisieren und protokollieren.

Aus dem klassischen Data-Warehouse-Umfeld hat sich **Talend Open Studio**<sup>7</sup> (TOS) in einer Reihe von Forschungsverbänden bewährt. Aus der tranSMART Foundation heraus wird in vielen Projekten **Kettle**<sup>8</sup> aus der Pentaho Data Integration and Business Analytics Platform zur Datenintegration verwendet.

In biomedizinischen Projekten wird häufig die **KNIME**<sup>9</sup> Analytics Platform verwendet, die eine sehr umfangreiche Funktionalität vom Lesen bis zum Versenden der transformierten Daten bietet. Speziell für die Anforderungen aus der medizinischen Verbundforschung ist komplementär das Integrated Data Repository Toolkit (IDRT) entstanden [12].<sup>10</sup> Das Toolkit bietet neben einer Installationshilfe zu i2b2 Importhilfen für viele Standardformate in der biomedizinischen Informatik und hilfreiche Werkzeuge für Terminologie- und Ontologie-Matching.

---

7 Siehe <https://de.talend.com>.

8 Siehe <http://community.pentaho.com/projects/data-integration/>.

9 Siehe <https://www.knime.org/knime>.

10 Integrated data repository toolkit (IDRT), siehe <http://community.i2b2.org/wiki/display/IDRT/>.

## Ontologie-Matching

Typischerweise werden Studiendaten zentralisiert gesammelt und verwaltet. Bei multizentrischen Forschungsprojekten werden die Daten zwar dezentral erfasst; ein Partner übernimmt aber in der Regel die Aufgaben des Datenmanagements und damit der Datenintegration für die Studie. Oftmals sind die Daten mit studienspezifischen Ontologien und Taxonomien verknüpft, die sich aus den verwendeten Untersuchungs- und Fragebögen (Case Report Forms, CRF) ergeben (z. B. LIFE Investigation Ontology [13, 14]). Die Übertragung in ein studienübergreifendes Forschungsdaten-Repository samt Datenintegration setzt die vorherige Integration der jeweilig in den Studien verwendeten Ontologien voraus. Dazu sind diese mit der im zentralen Forschungsdaten-Repository verwendeten Ontologie abzugleichen. Dieser Abgleich bezeichnet den Prozess des Ontologie-Matching und hat zum Ziel, semantisch gleichartige bzw. ausreichend ähnliche Ontologiekonzepte und damit verknüpfte Studien-Items der unterschiedlichen studienspezifischen CRFs einander zuzuordnen. Somit werden die Daten unter Nutzung einer einheitlichen Ontologie im Forschungsdaten-Repository abfragbar. Je präziser und umfassender die Daten der einzelnen Studien mit standardisierten Metadaten versehen sind, desto einfacher ist dieser Schritt umzusetzen.<sup>11</sup>

Ein schwerwiegendes Problem beim Abgleich phänotypischer klinischer Daten über Forschungsprojekte oder Studien hinweg besteht allerdings in der geringen terminologischen Standardisierung der Erfassung bzw. ihrer mangelhaften ontologischen Fundierung [15]. Für die Integration dieser Daten mit OMICS-Daten und die Weiterentwicklung einer individualisierten Medizin ist dieser Zustand allerdings als hoch problematisch anzusehen [16]. Vor diesem Hintergrund wird auch für die Nutzung einer Referenz-Terminologie wie SNOMED-CT geworben, die das Matching-Dilemma zumindest teilweise entschärfen könnte [15]. Eine Referenz-Terminologie könnte als Brücke zwischen den verschiedenen jeweils verwendeten und eher spezialisierten Terminologien und Klassifikationen dienen und würde zudem erlauben, unterschiedliche Datensätze mit Hilfe eines gemeinsamen Bezugssystems semantisch zu annotieren [17].

Neben der Nutzung im Studienkontext sind Ontologien in den Lebenswissenschaften auch zur Beschreibung (Annotation) von biologischen Objekten (Gene, Proteine etc.) weit verbreitet. So dient die Gene Ontology zur Annotation von Genprodukten; mit ihr werden molekulare Funktionen, die Beteiligung in biologischen Prozessen sowie die Lokation in zellulären Komponenten beschrieben. Weitere zahlreiche Ontologien in diesem Bereich werden von unterschiedlichen Institutionen zur Verfügung gestellt.

---

<sup>11</sup> Siehe hierzu das Kapitel zu Metadaten im vorliegenden Report.

Mit dem Ontology-Matching werden Korrespondenzen zwischen den Konzepten unterschiedlicher Ontologien erzeugt, sei es um inhaltliche Überlappungen zu kennzeichnen oder um die inhaltliche Vernetzung der Ontologien – und damit aus Sicht der Informatik den Zuwachs formalisierten Wissens – voranzutreiben.

Das Ontology-Matching ist seit einigen Jahren ein aktives Forschungsgebiet [18-21].<sup>12</sup> In den letzten Jahren wurde eine Reihe von Tools entwickelt, die eine semiautomatische Vorgehensweise unterstützen (z. B. [12] im Kontext von i2b2, für eine Evaluation anderer Tools vgl. [14, 20]). Viele davon sind jedoch Forschungsprototypen oder werden abseits der Medizininformatik verwendet, so dass sie sich in der medizinischen Studienpraxis erst noch bewähren müssen.

### **Datenhaltungssysteme / Forschungsdaten-Repositories**

Validierbare Datenhaltungssysteme sind in der biomedizinischen Forschung elementar, um beispielsweise den Bearbeitungsprozess (inkl. aller ausgeführter Operationen), ausgehend von den Rohdaten über ggf. mehrere Zwischenschritte bis hin zu den Ergebnisdaten unter Nutzung unterschiedlicher Werkzeuge, nachvollziehbar dokumentieren zu können.<sup>13</sup> Diese Prozessdokumentation ist nicht nur für die Verarbeitung der medizinischen Daten notwendig, sondern auch für die Gewinnung, Lagerung und Analyse von Bioproben. Zu jedem Aliquot muss erfasst werden, aus welcher Probe es gewonnen und in welcher Laboruntersuchung (z. B. zur Untersuchung von Biomarkern) es im späteren Verlauf verwendet wurde. Schließlich ist zu dokumentieren, welche Daten aus dieser Untersuchung resultieren und wie aus diesen interpretierbare Ergebnisdaten abgeleitet werden können, beispielsweise in Form einer Mikrobiom-Tabelle mit Spezies und Auftretenshäufigkeiten.

Neben der Verwendung von relationalen Datenbanksystemen sind in den letzten Jahren einige auf die Verwaltung von Forschungsdaten spezialisierte Systeme entwickelt worden. Dazu zählen openBIS [22] und iRODS. Das System **openBIS**<sup>14</sup> entstand im Umfeld der ETH in Zürich. Der funktionale Schwerpunkt liegt auf der feingranularen und metadatenannotierten Ablage biomedizinischer Daten und Methoden. Die Software wird derzeit an mehreren Standorten in Deutschland evaluiert bzw. bereits eingesetzt.

Das Integrated Rule-Oriented Data System (**iRODS**) entstand am RENCi in Chapel Hill in North Carolina mit einem primär dateibasierten Ansatz. iRODS<sup>15</sup> wird bran-

---

12 Siehe hierzu auch <http://ontologymatching.org>.

13 Siehe hierzu das Kapitel „Molekularbiologische Daten aus Hochdurchsatz-Analysen“ im vorliegenden Report.

14 Siehe <https://sis.id.ethz.ch/software/openbis.html>.

15 Siehe <http://irods.org>.

chenübergreifend für datenintensive akademische Anforderungen in derzeit mehr als 30 Projekten weltweit eingesetzt.

## Common Data Models

Common Data Models haben insbesondere im Kontext der PCORnet-Initiative (Patient Centered Outcome Research) in den USA stark an Bedeutung gewonnen. Das bereits in 2008 aufgestellte Mini-Sentinel-Programm der US Federal Drug Administration (FDA) für Pharmakovigilanzauswertungen [23] wird hierbei häufig als Ausgangspunkt und Vergleichsmaßstab für die in den nachfolgenden Projekten entwickelten CDMs angeführt [24]. Während es sich bei Mini Sentinel um ein geschlossenes System handelt, wurde im OMOP-Projekt (Observational Medical Outcomes Partnership) ein frei verwendbares CDM entwickelt [25] das im Nachfolgeprojekt OHDSI (Observational Health Data Sciences and Informatics) [26] konsequent fortgeführt wird und aktuell in seinem 5. Release vorliegt.<sup>16</sup> Das OMOP-Datenmodell enthält sowohl generische Tabellen, die (vergleichbar zu i2b2) verschiedenste Datenelemente aufnehmen können, als auch detaillierte strukturierte Tabellen für ausgewählte Konzepte (z. B. Medikation). Während es prinzipiell möglich ist, beliebige Vokabularien in eine OMOP-Datenbank zu importieren, sind eine Reihe von zentral unterstützten Standardterminologien vorgegeben, deren Nutzung für standortübergreifende Abfragen erforderlich ist. OHDSI stellt über das CDM hinaus eine Reihe von Werkzeugen zur Analyse und Auswertung der Daten bereit.<sup>17</sup> Das PCORnet-CDM stellt eine Weiterentwicklung von Mini Sentinel dar und wird von den verschiedenen innerhalb der PCORnet-Initiative aufgebauten Forschungsnetzen eingesetzt.<sup>18</sup> Es wird ebenfalls kontinuierlich weiterentwickelt und liegt aktuell im Release 3 vor. Auch hier ist die Nutzung vorgegebener Standardterminologien für standortübergreifende Abfragen erforderlich.

## Filter- und Analysewerkzeuge

Aus den NIH-geförderten Nationalen Zentren für biomedizinisches Computing in den USA ging das Werkzeug **i2b2**<sup>19</sup> (Informatics for Integrating Biology and the Bedside, [27]) hervor. Die i2b2-Plattform bietet über eine einfache Nutzeroberfläche in einer Baumstruktur die eingebundenen Datenquellen an. Aus diesen Datenquellen können Items für einfache und/oder-Verknüpfungen gewählt und als Filterkriterium verwendet werden. Als Ergebnis bekommt der Nutzer die Anzahl der Datensätze, die den Kriterien entsprechen. Optional können auch die Datensätze selbst exportiert werden.

---

16 Siehe <http://omop.org/CDM>.

17 Siehe <http://www.ohdsi.org/analytic-tools/>.

18 Siehe <http://www.pcornet.org/pcornet-common-data-model/>.

19 Siehe <https://i2b2.org/>.

i2b2 wird in den USA, aber auch in Europa sehr intensiv genutzt und weiterentwickelt. Die TMF begleitet das i2b2-Projekt seit vielen Jahren. In einem Cooperation Agreement zwischen dem i2b2 Center for Biomedical Computing in Boston und der TMF wurde eine engere Abstimmung und ein verstärkter Austausch zu den jeweiligen Aktivitäten in Bezug auf i2b2 vereinbart. Dies beinhaltet unter anderem die aktive gegenseitige Teilnahme an den jährlichen Treffen der Academic Users' Group in Boston bzw. in Europa.

**tranSMART**<sup>20</sup> [28] bietet neben den Funktionen von i2b2 eine intuitive Visualisierung des gewählten Datenbestandes (z. B. Altersverteilung, Geschlecht etc.) an. Zusätzlich können bereits nach einer Standard-Installation in tranSMART umfangreiche Analysewerkzeuge für Biomarkerdaten verwendet werden. Darüber hinaus können auch eigene R-Skripte in die Software und die Nutzeroberfläche eingebunden werden. Das Datenmodell von tranSMART fußt allerdings derzeit auf einem älteren Datenmodell von i2b2, was die Flexibilität bei komplizierteren Datenmodellen einschränkt. Die tranSMART Foundation wird derzeit u. a. von mehreren großen Pharmaunternehmen unterstützt. Als weitere Pluspunkte sind neben dieser Unterstützung – und möglicherweise auch eine Vorbedingung für diese – die präzise Roadmap-Planung und schnelle Releasezyklen hervorzuheben.

Bei den Treffen der i2b2 Academic Users' Group wurde die Dualität aus i2b2/tranSMART mehrfach kritisiert. Als Reaktion wurde eine engere Abstimmung der Releaseplanung beider Tools vereinbart und die Integration des aktuellen i2b2-Datenmodells in tranSMART angekündigt. Fortschritte zu dieser Arbeit wurden zuletzt beim europäischen Nutzertreffen in Italien<sup>21</sup> vorgestellt.

Mittlerweile werden am Markt mehrere kommerzielle Datenmanagementlösungen von der Extraktion der Daten bis hin zu kundenspezifischen Nutzeroberflächen für beide Plattformen angeboten, u. a. TrinetX, Biomera (beide i2b2) und The Hyve, Genialis (beide tranSMART).

Weitere Filter- und Analysewerkzeuge wurden (ohne Anspruch auf Vollständigkeit) in einem Review von Canuel et al. [10] behandelt. Daneben stellen auch kommerzielle Business-Intelligence-Lösungen immer mehr der hier vorgestellten Funktionen zur Verfügung (z. B. QlikView<sup>22</sup>).

---

20 Siehe <http://transmartfoundation.org/developers/>.

21 Siehe <http://cht.unipv.it/index.php/events/i2b2meeting2016-en>.

22 Siehe <http://www.qlik.com/products/qlikview>.

## Bewertung und Handlungsbedarf

Die ganze Verarbeitungskette von der Extraktion der verschiedenen Daten aus den jeweiligen Quellsystemen bis zur Analyse und Auswertung der Daten kann im biomedizinischen Umfeld mit der Kombination verschiedener Werkzeuge erreicht werden. Allerdings können viele der in der Verbundforschung hierfür eingesetzten Werkzeuge noch nicht die erforderliche Datenqualität und Reproduzierbarkeit der Analyseergebnisse garantieren. Für kritische Analysen schränkt dies derzeit die Auswahl geeigneter Werkzeuge noch deutlich ein. Im Zuge der Qualitätssicherung und Validierung der einzelnen Systeme in der Verarbeitungskette sind insbesondere auch die Schnittstellen zwischen den einzelnen Systemen im Sinne der guten wissenschaftlichen Praxis einer Qualitätssicherung zu unterziehen.

Neben der Bestimmung von Daten und Werkzeugen sind jedoch vor allem auch die Regeln festzulegen, nach denen Daten analysiert werden dürfen. Diese Policies müssen neben Datenschutz- und Intellectual-Property-Gesichtspunkten insbesondere auch berücksichtigen, wie Auswertungen im Sinne der guten wissenschaftlichen Praxis [29] qualitätsgesichert werden müssen. Neben einer gewissen Grundausbildung der beteiligten Forscher sind hier insbesondere Beratungsleistungen von biomedizinischer Informatik bzw. Biostatistik und Epidemiologie gefordert.

Auffällig ist, dass in vielen großen Forschungsfördermaßnahmen nicht auf bewährte Verfahren und Werkzeuge zurückgegriffen wird und stattdessen oft noch neue Entwicklungen von Datenmanagementwerkzeugen angestoßen werden, die nicht nachhaltig unterstützt und im erforderlichen Umfang qualitätsgesichert werden können. Hier ist im Sinne der Ökonomie und der guten wissenschaftlichen Praxis dringend eine bessere Abstimmung bzw. ein intensiverer Austausch zwischen Forschungsprojekten notwendig. Hierfür kann und sollte eine bewährte Plattform wie die AG IT-Infrastruktur und Qualitätsmanagement der TMF genutzt werden. Ein solcher forschungsbegleitender Austausch ist nicht nur für die Sichtung und Abstimmung generischer Architekturskizzen, sondern darüber hinaus auch zu den Themen „Access und Use-Policies“ oder „Betriebskonzepte für Kernmodule“ notwendig.

In dem hier soweit vorgelegten Kapitel zum Forschungsdatenmanagement müssen aus Ressourcengründen noch einige Themen unberücksichtigt bleiben, die erst in kommenden Versionen des IT-Reports in der nötigen Detailtiefe bearbeitet werden können. Hierzu gehören Persistent-Identifiers [vgl. 30], die automatisierte Erfassung und Berücksichtigung der Provenance von Datensätzen [31, 32] inklusive Metadaten zum Einwilligungstatus sowie die Verwendung geeigneter Terminologien bzw. die Refe-

renzung abgestimmter Data-Dictionaries. Einige dieser Themen sind beispielsweise im Technik-Report des DFG-Projektes „Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur“ (2013) angesprochen.<sup>23</sup>

Eine große Chance zur Professionalisierung und Verbesserung des Forschungsdatenmanagements an den klinischen Forschungsstandorten in Deutschland besteht in dem Förderkonzept Medizininformatik des BMBF.<sup>24</sup> Diese Chance kann aber nur sinnvoll genutzt werden, wenn das bestehende Know-how effektiv genutzt und unnötiges Wiederholen von Fehlern vermieden wird. Das Förderkonzept des BMBF sieht für die notwendigen Austausch- und Abstimmungsprozesse ein Nationales Steuerungsgremium samt Begleitprojekt vor. Es ist nicht nur zu wünschen, dass die hier dargelegten Erfahrungen effektiv von den aktuell geförderten Konsortien genutzt werden können, sondern darüber hinaus auch mit Hilfe der begleitenden Aktivitäten ein transparenter Rückfluss neuer methodischer Erfahrungen und Erkenntnisse in nachfolgende Versionen dieses Kapitels gewährleistet wird.

## Quellen und weiterführende Literatur

1. DFG *Leitlinien zum Umgang mit Forschungsdaten*. 2015. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf) (Abruf: 2016-02-05).
2. BMBF *Leitfaden zur Konzepterstellung für die Nationale Roadmap für Forschungsinfrastrukturen des Bundesministeriums für Bildung und Forschung*. 2015. Bundesministerium für Bildung und Forschung, [https://www.bmbf.de/pub/leitfaden\\_zur\\_konzepterstellung\\_forschungsinfrastruktur.pdf](https://www.bmbf.de/pub/leitfaden_zur_konzepterstellung_forschungsinfrastruktur.pdf) (Abruf: 2016-02-05).
3. Begley, C.G., *Six red flags for suspect work*. *Nature*, 2013. **497**(7450): S. 433 - 434.
4. RfII *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. 2016. Rat für Informationsinfrastrukturen, <http://www.rfii.de/?wpdmdl=1998> (Abruf: 2016-10-10).
5. DFG *Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Stellungnahme der Kommission für IT-Infrastruktur für 2016–2020*. 2016. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/foerderung/programme/wgi/kfr\\_stellungnahme\\_2016\\_2020.pdf](http://www.dfg.de/download/pdf/foerderung/programme/wgi/kfr_stellungnahme_2016_2020.pdf) (Abruf: 2016-11-23).

---

23 Siehe <http://www.forschungsdaten.org/index.php/Radieschen>.

24 Siehe <http://www.gesundheitsforschung-bmbf.de/de/medizininformatik.php>.

6. HRK *Empfehlung der 19. Mitgliederversammlung der HRK am 10. November 2015 in Kiel: Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können. Orientierungspfade, Handlungsoptionen, Szenarien.* 2015. Hochschulrektorenkonferenz, [http://www.hrk.de/uploads/tx\\_szconvention/Empfehlung\\_Forschungsdatenmanagement\\_10112015\\_01.pdf](http://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement_10112015_01.pdf).
7. Doods, J., Botteri, F., Dugas, M., Fritz, F., Ehr4Cr, W.P., *A European inventory of common electronic health record data elements for clinical trial feasibility.* *Trials*, 2014. **15**: S. 18.
8. Trinczek, B., Köpcke, F., Leusch, T. et al., *Design and multicentric Implementation of a generic Software Architecture for Patient Recruitment Systems re-using existing HIS tools and Routine Patient Data.* *Applied Clinical Informatics*, 2014. **5**(1): S. 264-283.
9. Christoph, J., Griebel, L., Leb, I. et al., *Secure Secondary Use of Clinical Data with Cloud-based NLP Services Towards a Highly Scalable Research Infrastructure.* *Methods of Informatics in Medicine*, 2015. **54**(3 2015): S. 276-282.
10. Canuel, V., Rance, B., Avillach, P., Degoulet, P., Burgun, A., *Translational research platforms integrating clinical and omics data: a review of publicly available solutions.* *Brief Bioinform*, 2014.
11. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0.* 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
12. Bauer, C.R.K.D., Ganslandt, T., Baum, B. et al., *Integrated Data Repository Toolkit (IDRT). A Suite of Programs to Facilitate Health Analytics on Heterogeneous Medical Data.* *Methods Inf Med*, 2016. **55**(2): S. 125-135.
13. Kirsten, T., Kiel, A. *Ontology-based Registration of Entities for Data Integration in large biomedical Research Projects. GI-Jahrestagung, Workshop - Informationsintegration in Service-Architekturen.* 2010. Leipzig, [http://www.btw-2015.de/res/proceedings/Workshops/DMS/Uciteli-Ontology-based\\_Retrieval\\_o.pdf](http://www.btw-2015.de/res/proceedings/Workshops/DMS/Uciteli-Ontology-based_Retrieval_o.pdf) (Abruf: 2016-02-08).
14. Uciteli, A., Kirsten, T., *Ontology-based retrieval of scientific data in LIFE, in Datenbanksysteme für Business, Technologie und Web (BTW 2015). Workshopband. 2.3.-3.3.2015, Hamburg* Hrsg.: N. Ritter, et al. 2015, Gesellschaft für Informatik, Bonn. S. 109-114.
15. Weil, P., Gusky, L., Bauer, C., Bockhorst, K., Semler, S.C., Rienhoff, O., *Thesenpapier des D-A-CH-Projekts. Empfehlungen an das Bundesministerium für Gesundheit, in Terminologien und Ordnungssysteme in der Medizin,* Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 7-20.

16. Rienhoff, O., *Die zunehmende Bedeutung von Terminologiesystemen in Forschung und Versorgung in einer individualisierten Medizin*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 29-34.
17. Dewenter, H., Thun, S., *Die SNOMED CT-Herausforderung*, in *Terminologien und Ordnungssysteme in der Medizin*, Hrsg.: O. Rienhoff and S.C. Semler. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin. S. 89-96.
18. Euzenat, J., Shvaiko, P., *Ontology Matching*. 2. Aufl. 2013, Springer, Heidelberg.
19. Bellahsene, Z., Bonifati, A., Rahm, E., Hrsg. *Schema Matching and Mapping*. 2011, Springer, Berlin Heidelberg.
20. Hartung, M., Groß, A., Kirsten, T., Rahm, E. *Effective Composition of Mappings for Matching Biomedical Ontologies. Extended Semantic Web Conference (ESWC) - Satellite Events, May 27-31, 2012. Revised Selected Papers*. 2012. Heraklion, Greece, [http://dbs.uni-leipzig.de/file/effectivemappingcomposition\\_extended\\_version\\_eswc.pdf](http://dbs.uni-leipzig.de/file/effectivemappingcomposition_extended_version_eswc.pdf) (Abruf: 2016-02-08).
21. Kirsten, T., Gross, A., Hartung, M., Rahm, E., *GOMMA: a component-based infrastructure for managing and analyzing life science ontologies and their evolution*. J Biomed Semantics, 2011. **2**: S. 6.
22. Bauch, A., Adamczyk, I., Buczek, P. et al., *openBIS: a flexible framework for managing and analyzing complex data in biology research*. BMC Bioinformatics, 2011. **12**: S. 468.
23. Forrow, S., Champion, D.M., Herrinton, L.J., Nair, V.P., Robb, M.A., Wilson, M., Platt, R., *The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program*. Pharmacoepidemiol Drug Saf, 2012. **21 Suppl 1**: S. 12-17.
24. Gini, R., Schuemie, M., Brown, J. et al., *Data Extraction and Management in Networks of Observational Health Care Databases for Scientific Research: A Comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE Strategies*. EGEMS (Wash DC), 2016. **4**(1): S. 1189.
25. Overhage, J.M., Ryan, P.B., Reich, C.G., Hartzema, A.G., Stang, P.E., *Validation of a common data model for active safety surveillance research*. J Am Med Inform Assoc, 2012. **19**(1): S. 54-60.
26. Hripcsak, G., Duke, J.D., Shah, N.H. et al., *Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers*. Stud Health Technol Inform, 2015. **216**: S. 574-578.
27. Kohane, I.S., Churchill, S.E., Murphy, S.N., *A translational engine at the national scale: informatics for integrating biology and the bedside*. J Am Med Inform Assoc, 2012. **19**(2): S. 181-185.

28. Athey, B.D., Braxenthaler, M., Haas, M., Guo, Y., *transSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research*. AMIA Jt Summits Transl Sci Proc, 2013. **2013**: S. 6-8.
29. DFG, *Memorandum: Proposals for Safeguarding Good Scientific Practice. Recommendations of the Commission on Professional Self Regulation in Science*. ergänzte Aufl. 2013, Wiley-VCH, Weinheim, [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf) (Abruf: 2016-11-21).
30. McMurry, J., Blomberg, N., Burdett, T. et al. *10 Simple rules for design, provision, and reuse of identifiers for web-based life science data*. 2015, <https://zenodo.org/record/31765> (Abruf: 2016-02-08).
31. Ragan, E.D., Endert, A., Sanyal, J., Chen, J., *Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes*. IEEE Trans Vis Comput Graph, 2016. **22**(1): S. 31-40.
32. Gil, Y., David, C.H., Demir, I. et al., *Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance*. Earth and Space Science, 2016. **3**(10): S. 388-415.



# Metadaten und Metadata Repositories

*Matthias Löbe*

## Anwendungsbereich

Die Verwertbarkeit von Primär- und Sekundärdaten der biomedizinischen Forschung hängt entscheidend von der Qualität und der Verfügbarkeit der sie beschreibenden Metadaten ab. Die Verwendung standardisierter Metadatenmodelle, -elemente und -kataloge fördert den Informationsaustausch zwischen IT-Systemen und sichert eine gemeinsame Interpretation der Daten.

In Deutschland werden jährlich hunderte klinische und epidemiologische Forschungsvorhaben (Register, Studien, Kohorten) gestartet. Die Notwendigkeit einheitlicher und detaillierter Definitionen von Datenelementen und Formularen ist dabei im Hinblick auf eine qualitativ hochwertige Dokumentation anerkannt. Der Aufwand für die Erstellung eines guten Dokumentationskonzepts ist hoch und erfordert die interdisziplinäre Zusammenarbeit von klinischen Forschern, Statistikern und IT-Experten. Gleiches gilt für Formulare und elektronische Masken in Informationssystemen der Versorgung.

In der Praxis werden Dokumentationskonzepte jedoch aufgrund fehlender Werkzeuge häufig ad hoc entworfen. Es existiert keine systematische, konsentrierte Standardmethode, wie sie beispielsweise Leiner und Haux in Form eines “documentation protocol” in Ergänzung zum Studienprotokoll fordern [1]. Fehler und Mehrdeutigkeiten führen zu Problemen bei der Qualität der erhobenen Daten (z. B. zu Queries in klinischen Studien) und können ganze Studiengenerationen betreffen. Andererseits ist die Ausarbeitung des Dokumentationskonzepts meist die Aufgabe jüngerer Forscher, die noch nicht über langjährige Erfahrung verfügen (z. B. welche Maßeinheiten und Referenzwerte in verschiedenen Laboren üblich sind). Ein weiteres Problem betrifft die mangelnde langfristige Pflege der meist papierbasierten Dokumentationskonzepte und deren Abgleich mit den Erfassungswerkzeugen oder Datenbanken im Falle von Änderungen an Merkmalen der Datenerhebung während des Vorhabens.

Eine konsequente Wiederverwendung bewährter Datenelemente würde prinzipiell eine schnellere und hochwertigere Spezifikation ermöglichen, d.h. die Startup-Kosten senken und den Studienstart beschleunigen. CDISC zitiert im Business Case for CDISC Standards [2] eine Untersuchung einer pharmazeutischen Firma, welche durch die Nutzung kontrollierter Metadaten und eines zentralen Metadatenregisters ein Einsparvolumen von 240.000 US-Dollar je klinischer Studie postuliert. Zusätzlich erleichtern gemein-

sam genutzte Dokumentationsmerkmale die Beschreibung und Zusammenführung von Datensätzen aus verschiedenen Quellen (Data Sharing) und deren übergreifende Auswertung (Metaanalysen). Die unabsichtliche Durchführung mehrfacher Forschungsvorhaben zur gleichen Fragestellung ist in Bezug auf die verwendeten Ressourcen unökonomisch und im Falle der Beteiligung menschlicher Probanden auch ethisch bedenklich.

Förderer aus dem öffentlichen Umfeld verlangen von Antragstellern daher immer häufiger die Erstellung und Durchsetzung eines Data Management and Sharing Plans (DMP) [3]. Der DMP legt typischerweise fest, welche Daten wie erhoben werden sollen, welche Maßnahmen zur Erhaltung des Wertes der Daten (z. B. Zugriffsmöglichkeit, Interpretierbarkeit, Erhalt von Zwischenergebnissen, Rechenwegen und Auswerteprogrammen) getroffen werden und welche Regeln und Restriktionen zum Teilen der Daten mit anderen Forschern definiert wurden. Auch die Deutsche Forschungsgemeinschaft verlangt inzwischen bereits bei Beantragung neuer Vorhaben eine Erklärung zum Forschungsdatenmanagement [4], da der Wert aufwendig erhobener oder erzeugter Primärdaten erkannt wurde [5]. Offensichtlich ist auch dies keine Aufgabe für einen einzelnen Wissenschaftler, sondern erfordert Kenntnisse aus klinischer Forschung, Biometrie, Medizinischer Informatik und Medizinischer Dokumentation.

Definitionen: Metadaten sind beschreibende Daten über Objekte einer Domäne. Datenelemente sind Kompositionen aus Metadaten zu einer funktionalen Einheit. Sie definieren Datenstrukturen zur konzeptuellen und repräsentativen Beschreibung der Bedingungen, unter denen einzelne Daten erhoben, berechnet oder gemessen wurden. Ein Metadata Repository (MDR) ist eine zentrale Datenbank oder ein Register von Datenelementen und ggfs. auch ein Werkzeug zur Verwaltung und zum Zugriff auf diese. Des Weiteren existieren im Bereich der Medizin viele begriffliche Ordnungssysteme (Terminologien), die als semantisches Fundament für Datenelemente dienen können, indem sie mit typisierten Relationen (Annotationen) verknüpft werden. Metadaten beschreiben demnach nicht nur Kategorien klinischer Fakten desselben Typs (Haarfarbe={blond, braun, rot, schwarz, ...}), sondern auch Kontextinformationen wie Erhebungsmethoden, Zugriffsrechte, Messgeräte oder Speicherorte. Bekanntestes Beispiel für einen Katalog wohldefinierter Metadaten sind die Dublin-Core-Elemente zur bibliografischen Beschreibung von Ressourcen.<sup>1</sup>

Eine Stärkung des wissenschaftlichen Interesses an Metadaten ist im Zuge der Bemühungen um die Entwicklung und Reifung von Technologien für ein Semantic Web zu beobachten. Ziel ist hier die Schaffung selbstbeschreibender Daten durch Fachvoka-

---

<sup>1</sup> Siehe <http://dublincore.org/documents/dces/>.

bularien oder Ontologien und deren maschinelle Verarbeitung durch Softwareagenten, die auf Verfahren der Künstlichen Intelligenz und der Logik zurückgreifen. Aufbauend auf den Technologien des WWW wurden in den letzten Jahren komplexe Repräsentations- (RDF, OWL), Abfrage- (SPARQL) und Regelsprachen (RIF) sowie spezielle Softwarekomponenten (Reasoner, Triple Stores) und Transportformate (JSON-LD) entwickelt. Obwohl sich diese Ideen im Bereich der Medizininformatik nur langsam durchsetzen – im Gegensatz zum Beispiel zur Bioinformatik, welche als „neue“ Disziplin nur wenige bestehende Vokabulare kannte und daher offener für neue Ansätze sein konnte – lässt sich gegenwärtig ein Trend zu einer parallelen Unterstützung verschiedener Datenmodelle feststellen, beispielsweise die Abbildung von CDISC Standards in XML und RDF.<sup>2</sup>

Das grundlegende Konzept des Semantic Web, einen einzigen globalen Graph per URIs adressierbarer Knoten zu schaffen, welche über HTTP angesprochen werden und strukturierte Informationen über sich und Verweise auf verwandte Knoten liefern (Linked Data), ist zweifellos elegant. Es erfordert von Datenanbietern eine hohe Qualität der Datenquellen sowie einen hohen Vernetzungsgrad und beseitigt gleichzeitig den Bedarf für lokale, redundante Kopien von Datenbeständen, welche häufig schlecht gepflegt werden. Beispielhaft seien die vielen veralteten Listen und Register klinischer Studien mit Zentren und Ansprechpartnern angesprochen, die sich in unzähligen Einrichtungen der klinischen Forschung finden (Abteilungen der Universitätskliniken, Studienzentren, KKS, Gesundheitszentren, IFBs, ...) und mühsam dutzendfach dieselben Informationen pflegen. Es wäre einfacher und sinnvoller, die Kerndaten im Deutschen Register Klinischer Studien (DRKS) zu speichern, erweiterte Informationen aber aus der Datenbank des jeweiligen Sponsors zu lesen sowie diese mit den Daten zu Prüfzentren und Prüffärzten aus der entsprechenden Datenbank des korrespondierenden Zentrums zu verknüpfen. Die Verantwortung für Richtigkeit und Aktualität der Daten obläge immer der ohnehin verantwortlichen Institution. Redundante Doppelerfassungen würden vermieden. Geeignete Studienzentren und klinische Experten könnten schneller recherchiert werden, ebenso Qualifikationsunterlagen z. B. für Prüffärzte<sup>3</sup> (FDA Form 1572 – „education, training, and experience that qualify the investigator as an expert in the clinical investigation“). Dem steht natürlich ein entsprechender (einmaliger) Aufwand für den Aufbau einer solchen Infrastruktur gegenüber.

---

2 Siehe <http://www.cdisc.org/standards/dataexchange>.

3 Siehe <http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Forms/UCM074728.pdf>.

## Betreiber, Anwender, Nutzer

Der Betrieb von Metadaten-Managementlösungen orientiert sich meist an zwei Szenarien. Zum einen handelt es sich um lokale Installationen in einer Organisation als Teilkomponente einer Softwareinfrastruktur, z. B. als Data-Dictionary-Modul in einem kommerziellen Clinical Data Management System (CDMS). In solchen Fällen liegt der Fokus im Allgemeinen auf der Unterstützung einer konkreten Teilaufgabe wie dem Erstellen von elektronischen CRF-Masken. Andere Ideen wie die semantische Fundierung mit medizinischen Terminologien sind nicht vorgesehen und aufgrund der starken Abkapselung des Gesamtsystems auch nicht hinzufügbare. Die Zielgruppe sind hier Datenbankentwickler, welche ein bereits ausgearbeitetes Dokumentationskonzept umsetzen. Eine im Rahmen eines TMF-Projekts durchgeführte Umfrage zeigt, dass diese Module, selbst wenn sie vorhanden sind, selten genutzt werden, da der Aufwand für das Heraussuchen der richtigen Datenelemente größer ist als deren Neuerstellung im System [6].

Die andere Variante sind dedizierte Metadatenregister, welche explizit die Unterstützung vieler Forschungsvorhaben in verschiedenen Organisationseinheiten zum Ziel haben. Sie sind insbesondere bei großen pharmazeutischen Firmen und im akademischen Bereich im anglo-amerikanischen Raum verbreitet. Beispielsweise hat die Firma Eli Lilly im Jahr 2010 ihr Metadata Repository mit über 20.000 Datenelementen veröffentlicht<sup>4</sup>, welches als Basis für alle klinischen Studien des Unternehmens genutzt wird. Ähnliche Register und Werkzeuge existieren auch bei Hoffmann-La Roche.<sup>5</sup> Das National Cancer Institute (NCI) in den USA betreibt das Cancer Data Standards Registry and Repository (caDSR)<sup>6</sup>, welches für die NCI-geförderten Krebsstudien verbindlich ist. Im Gegensatz zu den im vorigen Abschnitt beschriebenen Komponenten werden zentrale Metadatenregister schon beim Entwurf des Dokumentationskonzepts von Statistikern oder Datenmanagern eingesetzt. Des Weiteren handelt es sich hier um zentrale Dienste, die meist von dritten Applikationen über eine Schnittstelle angesprochen werden können.

Im letzteren Fall existieren in der für das Metadatenregister verantwortlichen Organisation Expertenteams, die Datenelemente kuratieren und harmonisieren, auf Konsistenz, Qualität und Standardkonformität achten und den Lebenszyklus der Einträge im Register verwalten. Das heißt im Besonderen, dass Dokumentationskonzepte keine beliebigen Datenelemente enthalten dürfen, sondern neue Datenelemente in ein formales Review-Verfahren eingebracht werden müssen. Dem inhärenten Aufwand hierfür steht

4 Siehe <https://web.archive.org/web/20110314074300/http://lillyodmlibrary.codeplex.com/> [aktuell nicht online verfügbar].

5 Siehe <http://www.tmf-ev.de/Desktopmodules/Bring2Mind/DMX/Download.aspx?EntryId=20484&PortalId=0>

6 Siehe <https://wiki.nci.nih.gov/display/caDSR/>.

ein Nutzen in Form standardisierter Auswerteverfahren und besserer Vergleichbarkeit der Ergebnisse verschiedener Forschungsprojekte entgegen. Insofern sind die Statistiker hier die primären Nutznießer standardisierter Metadaten.

In der akademischen Forschung in Deutschland sind solche Ideen allenfalls in Ansätzen zu beobachten. Oftmals ergeben sich Standard-Datenelemente weniger aus einem systematischen Review der Zieldomäne oder Konsensusprozessen, sondern einfach aus einer wiederholten Benutzung existierender Dokumentationskonzepte (gegebenenfalls mit Anpassung an das aktuelle Vorhaben).

## Rahmenbedingungen

### Rechtliche Rahmenbedingungen

Aus regulatoriver Sicht gibt es für Metadaten wenige einschlägige Richtlinien. Nur in einzelnen Fällen ist die Verwendung konkreter Metadatensätze gesetzlich vorgeschrieben, so z. B. in Deutschland der einheitliche onkologische Basisdatensatz der Arbeitsgemeinschaft Deutscher Tumorzentren (ADT) und der Gesellschaft der epidemiologischen Krebsregister e.V. (GEKID) [7] für die Krebsregister.

Im Bereich der Zulassungsstudien für neue Arzneimittel und Therapieansätze ist die U.S. Food and Drug Administration (FDA) eine der weltweit wichtigsten Behörden. Im Rahmen ihrer Zusammenarbeit mit CDISC werden die Standards Define-XML zur strukturierten Abbildung von Studienmetadaten (Visiten, CRFs, Items, Codelists) und das Study Data Tabulation Model (SDTM) zur inhaltlichen Beschreibung der Observationsdaten als Einreichungsformate akzeptiert. In der akademischen Forschung in Deutschland haben sich diese Standards noch nicht in der Breite durchsetzen können, es sind jedoch inzwischen hilfreiche Werkzeuge als Ergebnis von TMF-Projekten<sup>7,8</sup> verfügbar.

Eine aktuelle Entwicklung hin zu einer größeren Transparenz im Bereich klinischer Studien ist in der „Policy on publication of clinical data“ der Europäischen Arzneimittel-Agentur EMA [8] sichtbar. Diese plant, für alle Zulassungsstudien nicht nur umfangreiche Dokumente wie den Clinical Study Report für Dritte zugreifbar zu machen, sondern auch – unter noch nicht endgültig geklärten Verfahren zum Schutz persönlicher Identifikation und der Unternehmensgeheimnisse – die eigentlichen Patientendaten. Die TMF war in den Kommentierungsprozess der EMA zur Entwicklung

---

7 Siehe [http://www.tmf-ev.de/Themen/Projekte/V034\\_01\\_SDTM\\_Wandler.aspx](http://www.tmf-ev.de/Themen/Projekte/V034_01_SDTM_Wandler.aspx).

8 Siehe [http://www.tmf-ev.de/Themen/Projekte/V021\\_01\\_MAKS\\_1.aspx](http://www.tmf-ev.de/Themen/Projekte/V021_01_MAKS_1.aspx).

einer Policy bezüglich der Publikation von klinischen Studiendaten eingebunden<sup>9</sup> und hat sich für die freie Verfügbarkeit des Dokumentationsprozesses zumindest in Form der Annotated CRFs ausgesprochen. Diese Entwicklung zu mehr Transparenz ist nicht auf legislative Bemühungen beschränkt. So haben sich verschiedene Pharmafirmen zu der Initiative TransCelerate BioPharma Inc. zusammengeschlossen, die ebenfalls Clinical Data Transparency unterstützen möchte.<sup>10</sup> Auch die GMDS-Fachvertreter für Medizinische Informatik, Biometrie und Epidemiologie fordern „Open Metadata“, d. h. einen freien Zugang zu Formularen und Merkmalskatalogen im Gesundheitswesen generell, nicht beschränkt auf Studienformulare, sondern auch unter Einschluss der Strukturen elektronischer Patientenakten [9].

Vielfach unterschätzt wird zuletzt die Frage des geistigen Eigentums an Metadaten und den zugrundeliegenden Metadatenstandards. Viele basale Technologien wie XML oder auch die CDISC-Standards dürfen diskriminierungsfrei, d. h. von allen Anwendern und für alle Zwecke verwendet werden, auch wenn sie einem Copyright und einer bestimmten Lizenz unterliegen. Für einige andere Standards ist dagegen gar nicht klar, welche lizenzrechtlichen Bestimmungen anzuwenden sind. Allein die freie Verfügbarkeit im Internet oder als Supplement einer Publikation ist keine hinreichende rechtliche Grundlage, Datenelemente in eigene Projekte zu übernehmen. Dies stellt ein Problem für zentrale Metadata Repositories dar, aus denen sich Dritte eigene Studienspezifikationen zusammensetzen. Es ist daher sowohl für Anbieter wie auch für Anwender empfehlenswert, eine definierte Lizenz anzufügen, auch und gerade dann, wenn keine Einschränkung der weiteren Verwendung intendiert ist. Hierfür kann auf Lizenzen für freie Inhalte, z. B. die Creative Commons<sup>11</sup> zurückgegriffen werden.

## **Organisatorische Rahmenbedingungen**

Die medizinische Community (z. B. Studiengruppen, Fachgesellschaften, Gesundheitszentren) bemüht sich zunehmend, abgestimmte Metadatenkataloge zu entwickeln, die als Kerndatensatz, Basisdatensatz, Minimaler Datensatz, Harmonisierter Datensatz oder Common Data Elements (CDE) bezeichnet werden. Darunter versteht man eine Menge von Datenelementen, die entweder projektübergreifend verpflichtend erhoben werden oder zumindest als Vorzugsvariante im Fall einer Erhebung dienen. Beispiele hierfür sind Kerndatensätze in der Intensivmedizin, wie von den Fachgesellschaften DIVI [10] und ANZICS<sup>12</sup> definiert, oder zu seltenen Erkrankungen [11]. Eine PubMed-Suche nach

---

9 Siehe [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Overview\\_of\\_comments/2014/09/WC500174220.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Overview_of_comments/2014/09/WC500174220.pdf).

10 Siehe <http://www.transceleratebiopharmainc.com/initiatives/clinical-data-transparency/>.

11 Siehe <http://creativecommons.org/licenses/>.

12 Siehe <http://www.anzics.com.au/Pages/CORE/data-tools.aspx>.

„(common) data element(s)“ (ausschließlich im Titel) ergab 116 Arbeiten seit 2010, was die Bedeutung dieses Themas verdeutlicht.

Noch relativ neu sind die von CDISC definierten Standarddatenelemente für Dokumentationsbögen (Clinical Data Acquisition Standards Harmonization, CDASH<sup>13</sup>) wie auch die für einzelne Krankheiten definierten Datenelemente (Therapeutic Area Standards<sup>14</sup>). Aufgrund der an der Entwicklung beteiligten starken Kollaborationspartner aus Industrie und Behörden, besitzen sie aber das Potenzial, prägenden Einfluss auf das zugehörige Forschungsumfeld auszuüben.

Bei solchen Kerndatensätzen handelt es sich um konsentierete, harmonisierte Datenelemente einer Gruppe von Domänenexperten, welchen eine hohe Qualität zugebilligt werden kann. Dasselbe gilt für Datenelemente aus validierten Assessments wie beispielsweise der Glasgow Coma Scale (GCS) oder Scores wie APACHE. Im Gegensatz zu ihrer hohen inhaltlichen Qualität und Validität sowie ihrer methodischen Entwicklung ist die formale Repräsentation (Variablennamen, Kodierungen, Formate, Maßeinheiten) deutlich seltener definiert. Wichtig für eine dauerhaft hohe Datenqualität ist es ferner, Qualitätsstandards für das Datenmanagement festzulegen und zu kontrollieren [12].

## **Psychologische Rahmenbedingungen**

Während aussagekräftige und standardisierte Metadaten große Vorteile für das Messen und Vergleichen von Daten haben, existieren durchaus psychologische Gründe, die für einige Wissenschaftler gegen die Nutzung etablierter Metadaten-Standards und gegen die Veröffentlichung solcher Daten sprechen. Ein Hauptgrund ist natürlich ein fehlendes Verständnis für die Vorteile bei direkt wahrnehmbarem zusätzlichem Aufwand. Manchen Forschern mangelt es aber auch an der philanthropischen Grundeinstellung, Dokumentationskonzepte, die mit großem Aufwand erstellt wurden, kompensationslos zu teilen. Umfragen zeigen, dass allein die Pflicht zur Namensnennung der ursprünglichen Autoren die Bereitschaft zum Teilen stark erhöht [6]. In einigen Fällen wird auch die Sorge geäußert, dass Patienten gewisse Teile von klinischen Studien wie z. B. Risiko-Scores oder erwartete Nebenwirkungen ohne Erläuterung durch ärztliches Personal fehlinterpretieren können und die Rekrutierungsrate darunter leidet. Andere Gründe sind die Angst, sich durch Preisgabe detaillierter Methoden fachlicher Kritik auszusetzen oder andere Forscher zu ähnlichen, eventuell früher publizierten Experimenten anzuregen.

---

13 Siehe <http://www.cdisc.org/cdash>.

14 Siehe <http://www.cdisc.org/therapeutic>.

Zusätzlich ist auch von Hürden bei der bloßen Übernahme fremder Dokumentationskonzepte auszugehen. Manche Forscher fühlen sich ggf. in ihrer wissenschaftlichen Freiheit eingeschränkt oder wännen Innovationspotential verschenkt, wenn sie bereits etablierte und standardisiert beschriebene Datenelemente verwenden sollen. Der mit der Übernahme standardisierter Dokumentationskonzepte potentiell verbundene methodische und qualitative Gewinn wird dann möglicherweise unter- und der wissenschaftliche Wert der eigenen, noch nie im Studienalltag eingesetzten noch gar bewährten Kodiervorgabe überschätzt.

Keiner dieser Einwände ist vollkommen unbegründet und es müssen alle Risiken und Einschränkungen fallbasiert eingeschätzt und bewertet werden.

### **Finanzielle Rahmenbedingungen**

Die meisten Metadatenstandards wurden mit dem Ziel einer umfassenden Anwendung durch Dritte entwickelt und sind aus diesem Grund kostenlos verfü- und nutzbar. Bei elementaren Konzepten wie „Autor“ oder „Änderungsdatum“, selbst in spezifischer Komposition, wäre aber auch zu fragen, ob ihnen eine hinreichende Schöpfungshöhe zugebilligt werden kann.

Anders liegt der Fall bei komplexeren Instrumenten der Datenerhebung und –interpretation. So kann die Verwendung validierter Testbögen durchaus vierstellige Beträge kosten.<sup>15</sup> In einigen Fällen sind auch nur bestimmte Varianten kostenpflichtig. So ist beispielsweise die originale Version des Lebensqualitätfragebogens SF-36 der RAND Corporation kostenfrei, die deutsche Übersetzung aber geschützt.

Es ist Aufgabe der leitenden Wissenschaftler eines Forschungsprojekts, die rechtlich korrekte Verwendung der Datenerfassungsinstrumente sicherzustellen. Helfen können dabei Datenbanken wie die Health and Psychosocial Instruments (HaPI)<sup>16</sup> des Behavioral Measurement Database Services (BMDS), die detaillierte bibliografische Informationen für tausende Checklisten, Skalen, Surveys und Tests enthält.

### **Kategorisierung der IT-Landschaft**

Metadaten-Managementlösungen im Umfeld der patientenorientierten klinischen Forschung sind in den seltensten Fällen kommerzielle Systeme mit festen Entwicklungszyklen und Supportverträgen. Anders ist dies allenfalls, wenn sie Teil einer größeren Softwarelösung sind, z. B. als Data Dictionary eines CDMS oder als Formularmodul

---

15 Siehe <http://www.testzentrale.de/programm/wechsler-intelligence-scale-for-children-fourth-edition.html?catId=18>.

16 Siehe <http://www.bmdshapi.com/>.

eines EMR-Systems. Deutlich in der Überzahl sind Free- und Open-Source-Lösungen, wobei es sich häufig um Eigenentwicklungen im Rahmen wissenschaftlicher Forschungsprojekte oder um Entwicklungen von Verbänden und Initiativen des Gesundheitswesens handelt. Vertreter dieses Typs werden im Abschnitt „Aktueller Bestand“ näher vorgestellt. Standardprodukte mit einem nennenswerten Marktanteil oder einer größeren Zahl lokaler Installation sind dem Autor nicht bekannt.

## Aktueller Bestand

Metadaten-getriebene Ansätze sind in der biomedizinischen Forschung sehr populär, da sie flexibel an sich ändernde Abläufe und Regeln anpassbar sind. Gerade für Metadata Repositories existieren eine Unmenge konkurrierender, sich zum Teil aber auch ergänzender Standards und daraus folgend eine Vielzahl unterschiedlicher Implementierungen.

Der mit Abstand bedeutendste Standard ist die Norm ISO/IEC 11179-3: Registry meta-model and basic attributes [13]. Sie ist nicht auf die Domäne der Medizin beschränkt, wird aber aktiv von Wissenschaftlern aus der biomedizinischen Forschung propagiert und weiterentwickelt [14]. ISO 11179 erlaubt äußerst detaillierte Beschreibungen von Datenelementen und ihren Bestandteilen, aber auch von Annotationen durch Terminologien oder Ontologien sowie des Registrierungsprozesses von Datenelementen in einem Register. Projektspezifische Erweiterungen sind mittels eines Package-Konzepts möglich. Die ISO-Norm ist verglichen mit den später betrachteten Modellen am mächtigsten, dadurch jedoch auch aufwendiger zu implementieren.

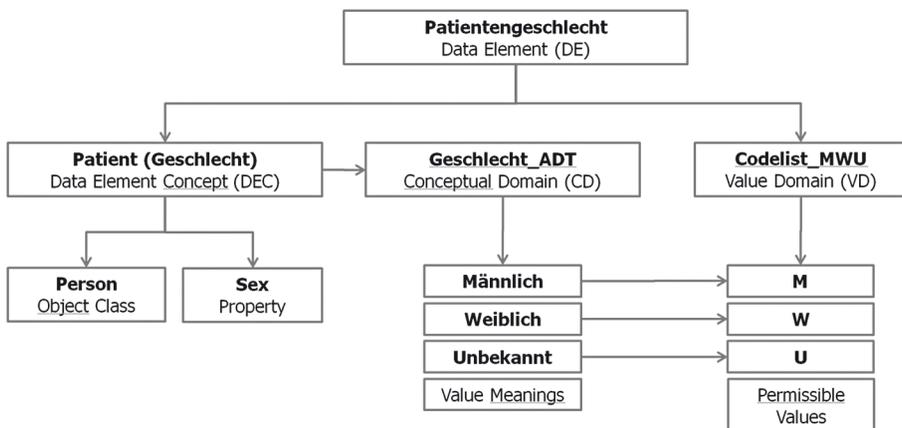


Abbildung 11.1: Stark vereinfachtes Beispiel für ein kategoriales Datenelement nach ISO 11179/3. Die wesentlichen Bestandteile sind: a) die inhaltliche Beschreibung (Data Element Concept), b) der inhaltliche Wertebereich (Conceptual Domain), beide mit Anknüpfungspunkten für Ontologien und Terminologien und c) die Repräsentation/Kodierung.

Aufgrund der oben genannten Vorteile bildet die ISO 11179 die Basis der meisten und der verbreitetsten MDRs. Besonders bekannt ist das bereit zuvor erwähnte caDSR [15, 16], welches neben der eigentlichen Datenbank eine vollständige Werkzeugunterstützung und eine API bietet. Das UK CancerGrid verfolgt einen ähnlichen Ansatz [17]. Dagegen ist die Australian Metadata Online Registry (METeOR)<sup>17</sup> ein webbasiertes System mit Fokus auf Metadaten der Gesundheitsversorgung, ebenso wie die United States Health Information Knowledgebase (USHIK)<sup>18</sup> oder das Data Dictionary des Canadian Institute for Health Information (CIHI)<sup>19</sup>.

PhEMA [18] ist ein derzeit in Entwicklung befindliches semantisches Metadata Repository zur Unterstützung des Phenotypings aus Electronic Health Records. Das EU-geförderte SALUS Project (Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies) entwickelt ein Metadata Repository namens SemanticMDR [19], in welches auch Datenelemente anderer EU-Projekte wie EHR4CR einfließen, und unterstützt das Profil Data Exchange (DEX) und weitere aus der Domäne Quality, Research and Public Health (QRPH) der Organisation Integrating the Healthcare Enterprise (IHE) [20]. Auch in Deutschland sind verschiedene Projekte auf Basis von ISO 11179 in Entwicklung [21, 22].

Die zukünftig größte Bedeutung könnte jedoch der CDISC Shared Health and Research Electronic Library (SHARE)<sup>20</sup> zukommen. SHARE beinhaltet die wichtigsten CDISC-Standards (CDASH, SDTM, SEND für nicht-klinische Daten wie Tierversuchsreihen und ADaM für statistische Analysen) in verschiedenen maschinenlesbaren Versionen: tabellarisch, als RDF/OWL und als ODM/XML (siehe dazu nächste Seite). Auch SHARE basiert in weiten Teilen auf der ISO-Norm 11179; einige Artefakte, z. B. *Metadata Element Sets* zur Gruppierung, wurden ergänzt. Als Referenzinformationsmodell dient das Biomedical Research Integrated Domain Group (BRIDG) Model<sup>21</sup>, welches seit 2015 als ISO 14199 vorliegt. Eine Web-API zum Zugriff aus externen Applikationen ist in Entwicklung. Aktuell ist SHARE nur für CDISC-Mitglieder zugänglich. Weiterhin in SHARE enthalten sind alle im Rahmen der CFAST-Initiative<sup>22</sup> entwickelten Kerndatensätze, die *Therapeutic Areas*. Jede Therapeutic Area (TA) deckt bestimmte Krankheiten wie Nierenkrebs oder Malaria ab und enthält typische und empfohlene Datenfelder zur Verwendung in klinischen Studien. Die Therapeutic Areas werden von CDISC mit verschiedenen Partnern, darunter die FDA, ausgearbeitet und durchlaufen einen

---

17 Siehe <http://meteor.aihw.gov.au/>.

18 Siehe <http://ushik.ahrq.gov>.

19 Siehe <http://www.cihi.ca/CIHI-ext-portal/internet/EN/TabbedContent/standards+and+data+submission/standards/>.

20 Siehe <http://www.cdisc.org/cdisc-share>.

21 Siehe <https://bridgmodel.nci.nih.gov>.

22 Siehe <https://c-path.org/programs/cfast/>.

formalen Entwurfs-, Qualitätssicherungs- und Freigabeprozess.<sup>23</sup> Ein zweiter Standard von internationaler Bedeutung ist die Clinical Document Architecture (CDA) [23] von Health Level 7 (HL7). Auch wenn derzeit keine CDA-basierten MDRs existieren, ist das CDA-Format inhaltlich aufgrund seiner engen Bindung an medizinische Terminologien und seiner expliziten Adressierung von Codes und Value Sets von Bedeutung. Eine praktische Relevanz ergibt sich aus der Verbreitung des Standards in Systemen der medizinischen Versorgung.

Fast Healthcare Interoperability Resources (FHIR)<sup>24</sup> [24] ist ein neuer Standardentwurf von HL7, der auf CDA und Nachrichten im Format HL7 Version 3 aufbaut, jedoch ein modernes Transportformat wie REST und leichter zu implementierende Datenstrukturen in JSON oder XML bzw. OAuth für die Authentifizierung verwendet. Für die ersten klinischen Systeme sind Schnittstellen prototypisch verfügbar, z. B. für ein MDR [18], ein Data Warehouse [25] oder genomische Apps [26].

Das Operational Data Model (ODM)<sup>25</sup> von CDISC ist ein wichtiges Austauschformat für Daten und Metadaten, wie sie typischerweise in klinischen Studien erhoben bzw. verwendet werden. Es ist XML-basiert, für viele Szenarien nutzbar [27] und kann auch andere klinische Datensätze gut repräsentieren [28]. Zudem lassen sich Wege zur Zusammenarbeit mit Systemen finden, die HL7 CDA unterstützen [29]. ODM findet zunehmend Verbreitung durch die Unterstützung in kommerziellen Studienmanagementsystemen [30]. Das Portal Medical Data Models [31] enthält eine der weltweit größten Sammlungen von ODM-Formularen (9.000 Stück) mit einem Schwerpunkt auf der klinischen Versorgung. Es wird aber verstärkt auch zu Aspekten der klinischen Forschung wie der Repräsentation von Ein- und Ausschlusskriterien, Secondary Use von EHR-Daten oder der Analyse von Häufigkeiten medizinischer Konzepte genutzt.

Parallel zu FHIR versucht eine weitere HL7-Arbeitsgruppe, die Clinical Information Modeling Initiative (CIMI)<sup>26</sup>, eine Sammlung von Datenmodellen originär aus dem Bereich der Versorgung zusammenzustellen, welche die semantische Interoperabilität in allen Bereichen des Gesundheitswesens unterstützen soll. Detailed Clinical Models (DCM) [32] sind Datenstrukturen, die durch formale Logiksprachen definiert werden und damit gut maschinenverarbeitbar sind. Konzeptuell besteht eine starke inhaltliche Nähe zu openEHR Archetypes [33], sodass beispielsweise openEHR-Softwarewerkzeuge verwendet werden können. DCM verwenden als Referenzterminologien SNOMED-CT

---

23 Siehe <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM371691.pdf>.

24 Siehe <http://hl7.org/fhir>.

25 Siehe <http://www.cdisc.org/odm>.

26 Siehe <http://www.opencimi.org>.

und LOINC. Derzeit existieren mehrere tausend kuratierte und verabschiedete DCMs, wobei Ergebnisse ähnlicher Ansätze wie Clinical Element Models (CEM) [34] darin eingegangen sind. Alle DCMs sind über ein frei zugängliches Repository verfügbar<sup>27</sup> und dürfen ebenso frei verwendet werden. Fernziel ist eine Verbindung zu FHIR über das Konzept der FHIR Dictionaries bzw. FHIR Data Elements.

Weitere relevante Standards und Initiativen sind das OMO Common Data Model (CDM) [35], die ISO-Norm 17369: Statistical Data and Metadata Exchange (SDMX) [36] oder der QUEXML-Standard der Data Documentation Initiative (DDI)<sup>28</sup> [37].

Abschließend sei auf Metadata Repositories verwiesen, die auf keinem publizierten Standard basieren, aber aufgrund ihrer Marktdurchdringung trotzdem eine einschlägige Quelle für Datenelemente sein können. Darunter fallen beispielsweise die Formularbibliotheken von Studienmanagementsystemen wie REDCap LOC [38], die OpenClinica CRF-Library<sup>29</sup> oder für epidemiologische Kohorten das MInCa Data Dictionary [39].

## Bewertung und Handlungsbedarf

Die theoretische Fundierung von Metadaten, -modellen und -systemen ist weit fortgeschritten. Es existieren vielfältige, reife, komplementäre (und konkurrierende) Ansätze für die unterschiedlichen Herausforderungen der patientenorientierten Forschung. Zu beobachten ist jedoch eine schwach ausgeprägte Verfügbarkeit reifer Werkzeuge, die den Forscher mit quantifizierbarem Mehrwert bei seinen Arbeitsprozessen unterstützen. Zu oft werden noch Insellösungen entwickelt, die nicht einmal den Anspruch erheben, mehr als eine projektspezifische Übergangslösung zu sein. Daraus resultieren starke Verzögerungen bei der Bereitstellung von Forschungsprimärdaten für Auswertungen und bei der Zusammenführung verschiedener Datensätze.

Auch die Standardisierung von Inhalten kommt nur langsam voran, denn die Definition und Harmonisierung von Datenelementen ist selten ein prioritäres Anliegen der biomedizinischen Forscher.<sup>30</sup> Vor diesem Hintergrund ist die Aussage der industrielastigen FHIR-Arbeitsgruppe durchaus positiv zu sehen, dass nur solche Elemente Eingang in den finalen (Kern-) Standard haben werden, die von ca. 80% der FHIR-unterstützten Systeme interpretiert und genutzt werden. Wünsche von Minoritäten sollen durch optionale Profile erfüllbar sein.

---

27 Siehe <http://www.opencem.org/>.

28 Siehe <http://www.ddialliance.org/tool/quexml>.

29 Siehe <https://library.openclinica.com/>.

30 Als frühes, positives Beispiel soll hier das im Dezember 2002 durchgeführte Forum "International Harmonisation of Trial Parameters for Malignant Lymphoma" des Kompetenznetz' Maligne Lymphome erwähnt werden (siehe <http://www.lymphome.de/Gruppen/AGQM/Projekte/IntHarm/index.jsp>), das eine internationale Zusammensetzung vorweisen und auf eine Werkzeugunterstützung zurückgreifen konnte.

Medizinische Terminologien und Begriffssysteme bilden seit längerem ein eigenes, etabliertes, nicht triviales Forschungs- und Anwendungsgebiet [40-42]. Ihr Beitrag zu Metadata Repositories besteht in der Verknüpfung von Datenelementen mit Konzepten aus Terminologien durch typisierte Annotationen. Im trivialen Fall soll damit eine semantische Äquivalenz ausgedrückt werden, es sind aber verschiedene schwächere Assoziationen denkbar. Die Zahl in Frage kommender Terminologien ist beträchtlich und hängt von Einsatzschwerpunkt und länderspezifischen Vorgaben ab. Sowohl die Einbettungsmöglichkeit wichtiger Terminologien wie MedDRA, LOINC oder SNOMED CT in ISO 11179 [43] als auch die Annotierung von Datenelementen in CDISC ODM mit Konzepten aus LOINC, SNOMED CT und UMLS [44] ist untersucht worden. Während die Verknüpfung von Datenelementen und Konzepten modelltechnisch machbar erscheint, ist die Aussage der Äquivalenz deutlich problematischer: Für Datenelemente kann es keine direkte Entsprechung [45], zu viele [46] oder keine vollständige [47, 48] geben. Praktische Studien zeigen des Weiteren, dass selbst bei Experten kein signifikantes Maß an Übereinstimmung in der Kodierung von Datenelementen auf Case Report Forms erzielt wurde [49]. Dedizierte Terminologie-Services wie BioPortal [50], LexGrid [51], SHARPN [52] oder CTS2-LE [53] können hier zukünftig helfen, bessere Ergebnisse zu erzielen. Weitere Fortschritte lässt die Reifung von Ontologien zur Beschreibung klinischer Konzepte erhoffen. Zu nennen wären die Data Element Ontology [54], OCRE [55] oder HDOT [56].

Neben den inhaltsbeschreibenden Metadaten werden *Verwaltungsmetadaten* (provenance data) zukünftig stark an Bedeutung gewinnen. Für viele wissenschaftliche Fragestellungen sind die Umstände der Datenerhebung (Einwilligungsstatus, Geräte, Verfahren, Chemikalien, Softwareversionen, Qualitätsmetriken, Eigentümerschaften) von großer Bedeutung für die Verwendung über den Primärzweck hinaus (Langzeitarchivierung, Reproduzierbarkeit, Meta-Analysen, Open/Shared Data). So hat der Rat für Informationsinfrastrukturen (RfII) in seinen „Empfehlungen zum Management von Forschungsdaten“ [57] im Hinblick auf eine zu schaffende Nationale Forschungsdaten-Infrastruktur eine neue „Datenkultur“ in dieser Richtung eingefordert.

## Literatur

1. Leiner, F., Haux, R., *Systematic planning of clinical documentation*. Methods of Informatics in Medicine, 1996. **35**(1): S. 25-34.
2. Vadakin, A., Kush, R. *Business Case for CDISC Standards, Stage V, 2014 Update*. 2014. CDISC, <http://www.cdisc.org/business-case> (Abruf: 2016-11-27).
3. Büttner, S., Hobohm, H.-C., *Handbuch Forschungsdatenmanagement*. 2011, Bock + Herchen, Bad Honnef.

4. Winkler-Nees, S. *Der Umgang mit Forschungsdaten in Wissenschaft und Lehre. Heisenberg-Vernetzungstreffen*. 2010. Bad Honnef: DFG, [http://www.dfg.de/download/pdf/dfg\\_magazin/wissenschaftliche\\_karriere/heisenberg\\_treffen\\_2010/forschungsdaten.pdf](http://www.dfg.de/download/pdf/dfg_magazin/wissenschaftliche_karriere/heisenberg_treffen_2010/forschungsdaten.pdf) (Abruf: 2016-11-27).
5. DFG *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*. 2009. Deutsche Forschungsgemeinschaft, Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, Unterausschuss für Informationsmanagement, [http://dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf) (Abruf: 2016-11-21).
6. Löbe, M. *Evaluation Metadata Repository (MDR-Eval). TMF-Projekt V063-02M. Abschlussbericht*. 2015. TMF, [https://www.tmf-ev.de/Themen/Projekte/V063\\_02M\\_Evaluation\\_MDR\\_Prototyp.aspx](https://www.tmf-ev.de/Themen/Projekte/V063_02M_Evaluation_MDR_Prototyp.aspx) (Abruf: 2016-11-27).
7. BMG *Bekanntmachung: Aktualisierter einheitlicher onkologischer Basisdatensatz der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID)*. 2014. Bundesministerium für Gesundheit, BAnz AT 28.04.2014 B2, [https://www.bundesanzeiger.de/ebanzwww/wexsservlet?page.navid=to\\_bookmark\\_officialsite&genericsearch\\_param.edition=BAnz+AT+28.04.2014](https://www.bundesanzeiger.de/ebanzwww/wexsservlet?page.navid=to_bookmark_officialsite&genericsearch_param.edition=BAnz+AT+28.04.2014) (Abruf: 2016-11-27).
8. EMA *Publication of clinical data for medicinal products for human use*. 2014. European Medicines Agency, POLICY/0070, [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Other/2014/10/WC500174796.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Other/2014/10/WC500174796.pdf) (Abruf: 2016-11-27).
9. Dugas, M., Jöckel, K.H., Friede, T. et al., *Memorandum "Open Metadata". Open Access to Documentation Forms and Item Catalogs in Healthcare*. *Methods of Information in Medicine*, 2015. **54**(4): S. 376-378.
10. Apfelstelt, F., Baldering, J.H., Bingold, T., *The DIVI and DGAI core data set ICM 2010*. *Anaesth Intensivmed*, 2010. **51**: S. 801-808.
11. Choquet, R., Maaroufi, M., de Carrara, A., Messiaen, C., Luigi, E., Landais, P., *A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research*. *J Am Med Inform Assoc*, 2015. **22**(1): S. 76-85.
12. Nonnemacher, M., Stausberg, J., Nasseh, D., *Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern*. 2. Aufl. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
13. ISO/IEC *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes*. 2013. International Organization for Standardization, ISO/IEC 11179-3:2013(E), [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=50340](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=50340) (Abruf: 2016-11-27).

14. Solbrig, H.R., *Metadata and the reintegration of clinical information: ISO 11179*. MD Comput, 2000. **17**(3): S. 25-28.
15. Nadkarni, P.M., Brandt, C.A., *The Common Data Elements for cancer research: remarks on functions and structure*. Methods Inf Med, 2006. **45**(6): S. 594-601.
16. Komatsoulis, G.A., Warzel, D.B., Hartel, F.W. et al., *caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability*. J Biomed Inform, 2008. **41**(1): S. 106-123.
17. Papatheodorou, I., Crichton, C., Morris, L. et al., *A metadata approach for clinical data management in translational genomics studies in breast cancer*. BMC Med Genomics, 2009. **2**: S. 66.
18. Jiang, G., Solbrig, H.R., Kiefer, R. et al., *A Standards-based Semantic Metadata Repository to Support EHR-driven Phenotype Authoring and Execution*. Stud Health Technol Inform, 2015. **216**: S. 1098.
19. Daniel, C., Sinaci, A., Ouagne, D. et al., *Standard-based EHR-enabled applications for clinical research and patient safety: CDISC - IHE QRPH - EHR4CR & SALUS collaboration*. AMIA Jt Summits Transl Sci Proc, 2014. **2014**: S. 19-25.
20. Sinaci, A.A., Laleci Erturkmen, G.B., *A federated semantic metadata registry framework for enabling interoperability across clinical research and care domains*. J Biomed Inform, 2013. **46**(5): S. 784-794.
21. Stausberg, J., Löbe, M., Verplancke, P., Drepper, J., Herre, H., Löffler, M., *Foundations of a metadata repository for databases of registers and trials*. Stud Health Technol Inform, 2009. **150**: S. 409-413.
22. Rimatzki, B., Haas, P., *ISO/IEC 11179 Metadata Repository als Basis für die Interoperabilität im Gesundheitswesen*, in *Informatik 2012, 42. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*, Hrsg.: U. Goltz, et al. 2012, GI, Braunschweig. S. 1275-1284.
23. Dolin, R.H., Alschuler, L., Beebe, C. et al., *The HL7 Clinical Document Architecture*. J Am Med Inform Assoc, 2001. **8**(6): S. 552-569.
24. Raths, D., *Trend: standards development. Catching FHIR. A new HL7 draft standard may boost web services development in healthcare*. Healthc Inform, 2014. **31**(2): S. 13, 16.
25. Schwartze, J., Jansen, L., Schrom, H., Wolf, K.H., Haux, R., Marschollek, M., *An HL7-FHIR-based Object Model for a Home-Centered Data Warehouse for Ambient Assisted Living Environments*. Stud Health Technol Inform, 2015. **216**: S. 1060.
26. Alterovitz, G., Warner, J., Zhang, P., Chen, Y., Ullman-Cullere, M., Kreda, D., Kohane, I.S., *SMART on FHIR Genomics: facilitating standardized clinico-genomic apps*. J Am Med Inform Assoc, 2015. **22**(6): S. 1173-1178.

27. Löbe, M., Aßmann, C., Beyer, R. et al., *Einsatzmöglichkeiten von CDISC ODM in der klinischen Forschung*, in *Informatik 2012, 42. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*, Hrsg.: U. Goltz, et al. 2012, GI, Braunschweig. S. 1295-1305.
28. Meineke, F.A., Stäubert, S., Löbe, M., Winter, A., *A comprehensive clinical research database based on CDISC ODM and i2b2*. *Stud Health Technol Inform*, 2014. **205**: S. 1115-1119.
29. El Fadly, A., Daniel, C., Bousquet, C., Dart, T., Lastic, P.Y., Degoulet, P., *Electronic Healthcare Record and clinical research in cardiovascular radiology. HL7 CDA and CDISC ODM interoperability*. *AMIA Annu Symp Proc*, 2007: S. 216-220.
30. Breil, B., Kenneweg, J., Fritz, F., Bruland, P., Doods, D., Trinczek, B., Dugas, M., *Multilingual Medical Data Models in ODM Format*. *Applied Clinical Informatics*, 2012. **3**(3): S. 276-289.
31. Dugas, M., Neuhaus, P., Meidt, A., Doods, J., Storck, M., Bruland, P., Varghese, J., *Portal of medical data models: information infrastructure for medical research and healthcare*. *Database*, 2016. **2016**.
32. Huff, S.M., Rocha, R.A., Coyle, J.F., Narus, S.P., *Integrating detailed clinical models into application development tools*. *Stud Health Technol Inform*, 2004. **107**(Pt 2): S. 1058-1062.
33. Garde, S., Knaup, P., Schuler, T., Hovenga, E., *Can openEHR Archetypes Empower Multi-Centre Clinical Research?* *Stud Health Technol Inform*, 2005. **116**: S. 971-976.
34. Jiang, G., Evans, J., Oniki, T.A. et al., *Harmonization of Detailed Clinical Models with Clinical Study Data Standards*. *Methods of Information in Medicine*, 2015. **54**(1): S. 65-74.
35. Overhage, J.M., Ryan, P.B., Reich, C.G., Hartzema, A.G., Stang, P.E., *Validation of a common data model for active safety surveillance research*. *J Am Med Inform Assoc*, 2012. **19**(1): S. 54-60.
36. *ISO/IEC Statistical data and metadata exchange (SDMX)*. 2013. ISO/IEC TC 154, [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=52500](http://www.iso.org/iso/catalogue_detail.htm?csnumber=52500) (Abruf: 2016-11-27).
37. Vardigan, M., Heus, P., Thomas, W., *Data Documentation Initiative: Toward a Standard for the Social Sciences*. *The International Journal of Digital Curation*, 2008. **3**(1): S. 107-113.
38. Obeid, J.S., McGraw, C.A., Minor, B.L. et al., *Procurement of shared data instruments for Research Electronic Data Capture (REDCap)*. *J Biomed Inform*, 2013. **46**(2): S. 259-265.

39. Meyer, J., Fredrich, D., Piegsa, J., Habes, M., van den Berg, N., Hoffmann, W., *A mobile and asynchronous electronic data capture system for epidemiologic studies*. Comput Methods Programs Biomed, 2013. **110**(3): S. 369-379.
40. Cimino, J.J., *Desiderata for controlled medical vocabularies in the twenty-first century*. Methods of Information in Medicine, 1998. **37**(4-5): S. 394-403.
41. Rector, A.L., *Clinical terminology: why is it so hard?* Methods Inf Med, 1999. **38**(4-5): S. 239-252.
42. Rienhoff, O., Semler, S.C., Hrsg. *Terminologien und Ordnungssysteme in der Medizin*. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
43. Ngouongo, S.M., Löbe, M., Stausberg, J., *The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research?* J Biomed Inform, 2013. **46**(2): S. 318-327.
44. Breil, B., Watermann, A., Haas, P., Dziuballe, P., Dugas, M., *Semantic enrichment of medical forms - semi-automated coding of ODM-elements via web services*. Stud Health Technol Inform, 2012. **180**: S. 1102-1104.
45. Pathak, J., Wang, J., Kashyap, S., Basford, M., Li, R., Masys, D.R., Chute, C.G., *Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience*. J Am Med Inform Assoc, 2011. **18**(4): S. 376-386.
46. Richesson, R.L., Fung, K.W., Krischer, J.P., *Heterogeneous but "standard" coding systems for adverse events: Issues in achieving interoperability between apples and oranges*. Contemp Clin Trials, 2008. **29**(5): S. 635-645.
47. Richesson, R.L., Andrews, J.E., Krischer, J.P., *Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research*. J Am Med Inform Assoc, 2006. **13**(5): S. 536-546.
48. Mougin, F., Burgun, A., Bodenreider, O., *Mapping data elements to terminological resources for integrating biomedical data sources*. BMC Bioinformatics, 2006. **7 Suppl 3**: S. S6.
49. Andrews, J.E., Richesson, R.L., Krischer, J., *Variation of SNOMED CT coding of clinical research concepts among coding experts*. J Am Med Inform Assoc, 2007. **14**(4): S. 497-506.
50. Salvadores, M., Alexander, P.R., Musen, M.A., Noy, N.F., *BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF*. Semant Web, 2013. **4**(3): S. 277-284.
51. Pathak, J., Solbrig, H.R., Buntrock, J.D., Johnson, T.M., Chute, C.G., *LexGrid: a framework for representing, storing, and querying biomedical terminologies from simple to sublime*. J Am Med Inform Assoc, 2009. **16**(3): S. 305-315.

52. Pathak, J., Bailey, K.R., Beebe, C.E. et al., *Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPN consortium*. J Am Med Inform Assoc, 2013. **20**(e2): S. e341-348.
53. Krebs, F. *Definition und Implementierung einer CTS2-standardisierten Abbildung von Terminologien aus dem Bereich des Infektionsschutzes*. 2013. Fachhochschule Brandenburg: Bachelorarbeit: Bachelorarbeit, <http://ots.fh-brandenburg.de/downloads/abschlussarbeiten/2013-09-02%20ba%20Franziska%20Krebs.pdf> (Abruf: 2016-11-27).
54. Jeong, S., Kim, H.H., Park, Y.R., Kim, J.H., *Clinical Data Element Ontology for Unified Indexing and Retrieval of Data Elements across Multiple Metadata Registries*. Healthc Inform Res, 2014. **20**(4): S. 295-303.
55. Ochs, C., Agrawal, A., Perl, Y. et al., *Deriving an abstraction network to support quality assurance in OCRE*. AMIA Annu Symp Proc, 2012. **2012**: S. 681-689.
56. Anguita, A., Garcia-Remesal, M., de la Iglesia, D., Graf, N., Maojo, V., *Toward a View-oriented Approach for Aligning RDF-based Biomedical Repositories*. Methods of Information in Medicine, 2015. **54**(1): S. 50-55.
57. RfII *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. 2016. Rat für Informationsinfrastrukturen, <http://www.rfii.de/?wpdmdl=1998> (Abruf: 2016-10-10).

# Identitätsmanagement

*Martin Lablans, Klaus Pommerening, Frank Ückert*

## Anwendungsbereich

Das Identitätsmanagement für Patienten oder Probanden in der medizinischen Forschung ist ein Kernthema des Datenschutzes und überall dort relevant, wo mit patientenbezogenen Daten geforscht wird. Identitätsmanagement dient dazu, die Identität der Betroffenen von den medizinischen „Nutzdaten“ zu trennen und durch die Zuweisung von Pseudonymen zu verbergen. Es erfüllt somit eine grundlegende Anforderung des Datenschutzes: Ein versehentlicher oder absichtlicher Missbrauch der Daten wird durch die Pseudonymisierung und gegebenenfalls den Einsatz eines Datentreuhänders wesentlich erschwert. Durch die verteilte Speicherung verschiedener Datenarten wird eine informationelle Gewaltenteilung erreicht.

Eine Besonderheit der medizinischen Verbundforschung ist, dass oft viele unabhängige Projekte kooperieren, die eigene Pseudonymisierungsschemata einsetzen müssen; daher muss das Identitätsmanagement auch in der Lage sein, verschiedene Pseudonyme zu verwalten und bei bestehender Notwendigkeit zusammenzuführen.

Soweit Forschung direkt im Behandlungszusammenhang stattfindet, ist der Patient persönlich bekannt und ein gesondertes Identitätsmanagement nicht nötig. Es wird aber sofort dann erforderlich, wenn für eine klinische Studie, ein Register oder eine Biobank eine örtlich oder zeitlich über die direkte Behandlung hinausreichende Datenspeicherung oder -verwendung vorgesehen ist.

Da Pseudonyme die Zuordnung zwischen medizinischen Daten und Identitätsdaten – unter der Verantwortung eines Datentreuhänders – bewahren, bleiben die Daten im rechtlichen Sinne personenbezogen; ihre Speicherung und Verarbeitung muss daher im Regelfall durch eine passende Einwilligungserklärung abgedeckt sein.

Besondere Probleme beim Identitätsmanagement entstehen bei internationalen Projekten, teils durch unterschiedliche Interpretation der rechtlichen Rahmenbedingungen, teils aber auch durch eine fehlende Internationalisierung der vorhandenen Werkzeuge.

Fazit: Das Identitätsmanagement ist vor allem für Forschungsverbünde (einschließlich deren Biobanken) ein unverzichtbares Mittel zur Erfüllung der Datenschutzanforderungen. Es ist daher auch eine zentrale Komponente des TMF-Leitfadens zum

Datenschutz [1-3]. Durch den Einsatz geeigneter Konzepte und Werkzeuge wird die Datenhaltung im Verbund rechtlich abgesichert, wobei die Bewertung der Maßnahmen oft von der zuständigen Datenschutzaufsicht abhängt. Hier hat die TMF bereits seit ihrem Beginn Pionierarbeit geleistet, indem sie in Abstimmung mit den Datenschutzbeauftragten des Bundes und der Länder mit den generischen Konzepten zuverlässige Modelle aufgezeigt hat, die als Vergleich oder Blaupause für neue Forschungsverbände dienen können.

## Betreiber, Anwender, Nutzer

Das Identitätsmanagement ist in erster Linie Teil der Infrastruktur eines Forschungsverbunds. Zur Gewährleistung der informationellen Gewaltenteilung wird es in der Regel als getrennter Netzdienst betrieben, wie es z. B. auch der PID-Generator<sup>1</sup> und der Pseudonymisierungsdienst<sup>2</sup> der TMF sowie als Weiterentwicklung die Mainzliste<sup>3</sup> unterstützen. Die eigentliche Nutzung der Dienste erfolgt in der Regel implizit, etwa weil Funktionen zum Ansprechen des Identitätsmanagements in Software zum Electronic Data Capturing (EDC) eingebaut sind. Hierfür kann z. B. die gemeinsam vom Universitätsklinikum Münster und der TMF zur Verfügung gestellte Datenschutz-Bibliothek<sup>4</sup> zum Klinischen Modul des TMF-Datenschutzleitfadens [1] (vgl. Modell A der ersten Version der generischen Datenschutzkonzepte der TMF [2]) oder ihre Weiterentwicklung, der Mainzliste.Client [4], genutzt werden. Direkte Anwender sind hingegen nur Verbundpartner mit Treuhänderfunktion, wobei die direkte Anwendung bei diesen sich im Regelfall auf Depseudonymisierungsprozesse und die Qualitätssicherung im Record Linkage beschränkt.

Indirekt sind alle Mitarbeiter eines Forschungsverbunds Anwender des Identitätsmanagements, indem sie die Infrastruktur nutzen. Der gesamte Forschungsverbund profitiert von einer datenschutzgerechten IT-Landschaft und von reduzierten IT-Investitionen durch den Einsatz vorhandener Werkzeuge und Konzepte, und sei es nur als Vorlage für Eigenentwicklungen.

Betreiber des Identitätsmanagements soll eine unabhängige vertrauenswürdige Instanz (TTP = Trusted Third Party oder Datentreuhänderdienst) sein. Wünschenswert ist eine Professionalisierung und Verstetigung der Dienste, sowohl in organisatorischer als auch in technischer Hinsicht. Modelle dafür kristallisieren sich erst heraus; bisher werden

---

1 Siehe <http://www.tmf-ev.de/Produkte/P015011>.

2 Siehe <http://www.tmf-ev.de/Produkte/P000011>.

3 Siehe <http://www.mainzliste.de>.

4 Siehe <http://www.tmf-ev.de/Produkte/P014012>.

die Dienste fast durchweg in Eigenregie der Forschungsverbände betrieben und die vorhandenen Konzepte oft eher zur Orientierung für eigene Implementierungen genutzt.

Eine wesentliche Anforderung an den Datentreuhänderdienst ist seine organisatorische Unabhängigkeit von den übrigen Teilnehmern des Forschungsverbands; das bedeutet eine Bindung nur an die Regeln des Verbunds und schließt jegliche Weisungsabhängigkeit aus. Insbesondere darf der Datentreuhänderdienst nicht als Auftragsdatenverarbeitung konzipiert werden [5]. Dabei kann die rechtliche Verantwortung vom technischen Betrieb getrennt werden, der seinerseits durchaus als Datenverarbeitung im Auftrag des Datentreuhänderdienstes implementiert werden kann. Ist ein solcher technischer Dienstleister für mehrere unterschiedliche Datentreuhänderdienste tätig, so muss er eine strikte Mandantentrennung für diese umsetzen, die den Anforderungen der Orientierungshilfe „Mandantenfähigkeit“ der Datenschutzbeauftragten [6] genügt.

## Rahmenbedingungen

Die wesentlichen **rechtlichen Rahmenbedingungen** für das Identitätsmanagement liegen im Datenschutzrecht und im Arztrecht (das zusammen mit dem Strafrecht die ärztliche Schweigepflicht festschreibt). Das Datenschutzrecht ermöglicht im Regelfall auf Basis einer informierten Einwilligung die rechtlich einwandfreie Nutzung von Daten für die Forschung. Das ebenfalls im Datenschutzrecht verankerte Gebot der Datensparsamkeit schreibt dabei immer eine Anonymisierung oder Pseudonymisierung der Daten vor, wenn dies den Forschungszweck nicht verhindert. Der Pseudonymisierung ist dabei häufig der Vorzug zu geben, da nur sie die langfristige Zusammenführung der Daten eines Patienten z. B. bei Follow-ups sowie die Rekontaktierung von Patienten zu Rekrutierungszwecken oder auch bei Vorliegen behandlungsrelevanter Forschungsergebnisse ermöglicht. Zudem erfordert eine effektive Anonymisierung häufig ebenfalls eine Vergrößerung der medizinischen Daten, die die weitere Nutzung für bestimmte Fragestellungen ausschließt.<sup>5</sup> Auch die im April 2016 verabschiedete und ab Mai 2018 geltende EU-Datenschutzverordnung (DSGVO) gibt den Weg eines pseudonymisierten Identitätsmanagements vor.

Die DSGVO formuliert den Begriff der Pseudonymisierung im Vergleich zum bisherigen Bundesdatenschutzgesetz strenger. Inwieweit dadurch andere als rein terminologische Anpassungen des TMF-Datenschutzleitfadens oder bestehender Datenschutzkonzepte notwendig werden, ist im Laufe der Anpassung der nationalen Datenschutzgesetzgebung und ihrer Interpretation zu prüfen.

---

<sup>5</sup> Siehe hierzu das Kapitel „Anonymisierung von medizinischen Individualdaten“ in diesem Report.

Eine wichtige **organisatorische Rahmenbedingung** ist durch die Erarbeitung des TMF-Datenschutzleitfadens [1] und dessen Abstimmung mit der Konferenz der Datenschutzbeauftragten des Bundes und der Länder<sup>6</sup> gegeben. Hier liegen generische Musterlösungen für medizinische Forschungsverbände vor, die schon in vielen Fällen umgesetzt wurden und zu einem akzeptierten und mit Datenschützern abgestimmten Datenschutzkonzept des jeweiligen Forschungsverbands führten.

Für den verteilten Betrieb der im Datenschutzleitfaden vorgegebenen Dienste des Identitätsmanagements werden geeignete unabhängige Dienstleister benötigt, z. B. medizininformatische Institute oder Rechenzentren. Hier sind in vielen Fällen bereits praktikable Lösungen geschaffen worden, für die aber noch ein rechtlich abgesicherter standardisierter Rahmen aufgebaut werden muss; eine wichtige Rolle spielen dabei Verträge, die insbesondere die nötige rechtliche Unabhängigkeit gewährleisten (Bindung an die Richtlinien des Verbundes, aber Weisungsunabhängigkeit), Policies und SOPs.

**Technische Rahmenbedingungen** wurden von der TMF schon frühzeitig durch universelle Softwarelösungen geschaffen, die z. T. schnell und einfach nutzbar sind, aber dringend einer weiteren Anpassung an aktuelle Systemumgebungen und einer Funktionserweiterung bedürfen. Die Installation auf dedizierten Servern in hochsicherer Umgebung ist in vielen Fällen verwirklicht. Hinsichtlich der Integration dieser für die deutsche Forschungslandschaft spezifischen Dienste fällt auf, dass standardisierte Schnittstellen für den Austausch von Identitätsdaten und zugehörigen IDs fehlen oder für eher versorgungsorientierte Anwendungsfälle konzipiert wurden, wie z. B. im Rahmen von HL7-Nachrichten oder in den IHE-Profilen PIX/PDQ.

Zu nennen sind hier auch **psychologische Rahmenbedingungen**. Insbesondere in neu gegründeten Forschungsverbänden besteht oft ein mangelndes Problembewusstsein hinsichtlich der notwendigen Datenschutzmaßnahmen und des hierfür erforderlichen Identitätsmanagements; dieses erfordert Aufklärungsarbeit seitens der TMF, wobei die Beweggründe und Erfahrungshintergründe der Beteiligten berücksichtigt werden müssen. Dieser dauerhaften Weiterbildungsaufgabe widmet sich die TMF mit vielen Workshops, den Jahreskongressen, der TMF-School und vor allem auch mit dem niedrigschwelligen Beratungsangebot der AG Datenschutz.<sup>7</sup> Den Verantwortlichen für die Verbände muss vermittelt werden, dass erst ein sorgfältig geplantes und sicher implementiertes Identitätsmanagement eine datenschutzgerechte Zusammenarbeit im

---

6 Vgl. Webseiten der Landesdatenschutzbeauftragten aus Hessen, Sachsen-Anhalt, Thüringen und Berlin. Weblinks unter [www.tmf-ev.de/news/1518](http://www.tmf-ev.de/news/1518).

7 Siehe [https://www.tmf-ev.de/Arbeitsgruppen\\_Foren/AGDS/Beratung.aspx](https://www.tmf-ev.de/Arbeitsgruppen_Foren/AGDS/Beratung.aspx).

Verbund über institutionelle Grenzen hinweg ermöglicht. Es stellt keine zusätzliche bürokratische Hürde dar, sondern hilft dabei, die Ziele des Verbunds zu erreichen. Die Angst vor der Komplexität und vor Umsetzungsschwierigkeiten kann durch die Angebote der TMF wesentlich reduziert werden.

Fazit: Das Konzeptstadium zur Erfüllung der rechtlichen Rahmenbedingungen kann heute mit generischen Vorlagen und einem breiten Beratungsangebot gut unterstützt werden, an professionellen Angeboten zur Umsetzung von technischen und organisatorischen Maßnahmen wird gearbeitet, das Problembewusstsein, insbesondere bei neuen Forschungsverbänden, ist immer wieder neu zu schaffen und zu schärfen.

## Kategorisierung der IT-Landschaft

IT-Systeme, die in der Verbundforschung an das Identitätsmanagement angebunden werden müssen, fallen in zwei Klassen:

1. IT-Systeme, die speziell für den Forschungsverbund beschafft oder geschaffen werden.
2. IT-Systeme, die der Routineversorgung dienen und Daten für den Forschungsverbund exportieren, sowie die dazu genutzten Exportschnittstellen.

In die erste Klasse gehören EDC-Systeme und Registersoftware sowie „selbst gestrickte“ Datenbanken und Datensammlungen bis hin zur Nutzung von Access-Datenbanken und Excel-Tabellen. Da oft Weboberflächen eingesetzt werden, ist die Einbindung des Identitätsmanagements mittels Webtechnologien besonders relevant. Auf der Infrastrukturebene muss das Identitätsmanagement in die genutzte Kommunikationssoftware eingebunden sein.

In die zweite Klasse gehören Kliniksysteme wie KIS, LIMS, RIS (aber auch APIS) sowie klinische Datawarehouses und Register samt Werkzeugen für Datenextraktion und Datenaufbereitung<sup>8</sup>.

Fazit: Damit das Identitätsmanagement in heterogene IT-Landschaften integriert werden kann, sollte die Implementation auf zentralen Diensten beruhen, die als Web-Services über einheitliche Schnittstellen genutzt werden können.

---

<sup>8</sup> Siehe die Abschnitte zur IT-Landschaft in den Kapiteln zur Sekundärnutzung und zum Forschungsdatenmanagement.

## Aktueller Bestand

Zurzeit werden in medizinischen Forschungsverbänden verschiedene organisatorische **Ansätze** verfolgt und verschiedene **Werkzeuge** zum Identitätsmanagement eingesetzt:

- ▶ Eine genaue Umsetzung des Identitätsmanagements, wie im Datenschutzleitfaden der TMF beschrieben, ist für das Klinische Modul recht häufig anzutreffen, zunehmend auch für das Forschungsmodul.
- ▶ Der PID-Generator der TMF wird in Verbänden, die ein Klinisches Modul umsetzen, an vielen Stellen eingesetzt, oft direkt, oft aber auch in modifizierter oder „nachgebauter“ Form.
- ▶ Der Pseudonymisierungsdienst der TMF, der für das Forschungsmodul benötigt wird, wird bisher nur experimentell genutzt; eine voll funktionsfähige echte Nutzung steht noch aus.
- ▶ Darüber hinaus werden in einigen Verbänden selbst entwickelte Lösungen genutzt, die dem TMF-Ansatz nachgebildet sind (und in einigen Fällen schon vor dem TMF-Konzept existierten).
- ▶ Von der TMF unabhängige Lösungen für das Identitätsmanagement findet man vor allem bei den epidemiologischen Krebsregistern. Darüber hinaus gibt es auch in anderen Projekten konzipierte und umgesetzte Ansätze (QuaSi-Niere, ACGT, GANI\_MED). Im TMF-Projekt zur Erstellung des Datenschutzleitfadens wurde eine Übersicht erarbeitet, und die interessantesten der einzelnen Ansätze wurden begutachtet. Insgesamt wurden sie als gute Ansätze, aber für einen generischen Forschungsverbund zum unmittelbaren Einsatz als nicht flexibel und leistungsfähig genug bewertet.<sup>9</sup>
- ▶ Die im Mai 2013 veröffentlichte „Mainzliste“ kann als geeignete Weiterentwicklung des PID-Generators angesehen werden. Das Werkzeug wird inzwischen durch mindestens 14 Verbundprojekte, Biobanken und Softwareimplementierungen verwendet.<sup>10</sup> An der Open-Source-Software arbeiten zunehmend auch Entwickler anderer Institutionen mit.
- ▶ Im Zuge der Entwicklung der Mainzliste wurde eine gleichnamige Schnittstelle auf Basis aktueller Webtechnologien entwickelt und separat unter Creative-Commons-Lizenz veröffentlicht [7]. Sie ist auf eine Nutzung in webbasierten Systemen zugeschnitten und wird mittlerweile in ersten – gleichermaßen universitären und kommerziellen – EDC-Systemen, Registern und Biobankverwaltungen implementiert.

---

9 Berücksichtigt wurden u. a. die Ergebnisse eines TMF-Workshops zu ID-Tools vom 22.5.2012 in Mainz und öffentlich verfügbare Unterlagen im Oktober 2012.

10 Vgl. Liste unter <https://bitbucket.org/medicalinformatics/mainzliste>.

- ▶ Ein ähnliches Leistungsspektrum bieten die in Greifswald im Rahmen des MO-SAIC-Projekts [8] veröffentlichten Werkzeuge E-PIX (zum Identitätsmanagement) und gPAS (zur Pseudonymverwaltung), die neben ihrer Unterstützung von IHE PIX/PDQ im Rahmen der NAKO Gesundheitsstudie auch das Kommunikationsprinzip der Mainzliste adaptiert haben [9]. Sie können als Teil der MOSAIC-Werkzeuge nach einmaliger Registrierung kostenfrei unter einer Open-Source-Lizenz bezogen werden.<sup>11</sup>
- ▶ Viele Verbände betreiben gar kein wirksames Identitätsmanagement.

Für das Identitätsmanagement gibt es verschiedene **Betriebsmodelle** von der „Pseudonymisierung an der Quelle“ bis zum einrichtungsübergreifenden Angebot zentraler Dienste. In Zukunft wird auch ein verbundübergreifendes Identitätsmanagement eine zunehmend wichtige Rolle spielen.

Das Angebot der TMF und einzelner Mitgliedsverbände in Bezug auf das Identitätsmanagement umfasst<sup>12</sup>:

- ▶ PID-Generator mit Patientenliste, fehlertolerantem Record-Linkage, deutscher Phonetik nach Michael [10] und Postel [11] sowie SOAP-Schnittstelle.
- ▶ Pseudonymisierungsdienst für die zweite Stufe einer doppelten Pseudonymisierung mit je einer Schnittstelle zu Studiendatenbanken und Forschungsdatenbanken,
- ▶ Interoperabilität mit EDC-Software – etwa in secuTrial (proprietärer PID-Dispatcher oder Mainzliste-Schnittstelle) oder generisch in Form der frei verfügbaren Datenschutzbibliothek zur Umsetzung von Modell A der ersten generischen Datenschutzkonzepte der TMF (jetzt integriert in das Klinische Modul des Datenschutzleitfadens der TMF) für javascriptfähige Browser bzw. mit Hilfe des Mainzliste.Client [4]. SecuTrial wird in dieser Form z. B. in den Kompetenznetzen „Parkinson“ und „Angeborene Herzfehler“ genutzt. Für den Einsatz der Datenschutzbibliothek zu „Modell A“ kann das Register des Netzwerks für Ichthyosen und verwandte Verhornungsstörungen (NIRK) als Referenz dienen.
- ▶ Direkte Begleitung und Einflussnahme auf die Fortentwicklung von IT-Werkzeugen ist im Rahmen der TMF-Arbeitsgruppen „IT-Infrastruktur und Qualitätsmanagement“ (IT-QM) und „Datenschutz“ möglich, beispielsweise im MAGIC-Projekt (s. u.)

Die in Greifswald (E-PIX und gPAS) und Mainz (Mainzliste) verfügbaren Werkzeuge werden auch als webbasierte Dienstleistungen angeboten, im letzteren Fall auch betrieben durch dritte, z. T. kommerzielle, Anbieter.

---

<sup>11</sup> Siehe <https://mosaic-greifswald.de/werkzeuge-und-vorlagen.html>.

<sup>12</sup> Siehe [www.tmf-ev.de/produkte](http://www.tmf-ev.de/produkte).

## Bewertung und Handlungsbedarf

Ein sorgfältig konzipiertes und sicher umgesetztes Identitätsmanagement ist für einen datenschutzgerechten Betrieb eines Forschungsnetzes unerlässlich. Es spielt eine zentrale Rolle im Datenschutzleitfaden der TMF und stellt dort eine zentrale Komponente eines Forschungsverbunds dar. Für die von der TMF oder ihren Mitgliedsverbänden angebotenen Werkzeuge kann man resümieren:

- ▶ Der PID-Generator wird in vielen Verbänden eingesetzt und hat seine Nützlichkeit und Stabilität bewiesen. Für neu beginnende Forschungsverbände sind die Mainzliste oder E-PIX/gPAS als Weiterentwicklungen geeignet und haben sich bereits bewährt.
- ▶ Der Pseudonymisierungsdienst der TMF wird derzeit noch nicht produktiv eingesetzt.
- ▶ Der Mainzliste.Client sowie die Software-Bibliothek zur Umsetzung des generischen Modells A der ersten TMF-Datenschutzkonzepte wird in wenigen Verbänden im Rahmen spezifisch entwickelter Anwendungen genutzt.

Um die Anwendung in der deutschen medizinischen Forschungslandschaft und die internationale Kooperation zu fördern, entwickelt der DFG-geförderte TMF-Mitgliedsverbund MAGIC<sup>13</sup> IT-Werkzeuge und Prozesse weiter, sodass sie auch komplexe Anforderungen des TMF-Datenschutzleitfadens umsetzen können:

- ▶ Die Funktionen des Pseudonymisierungsdienstes der TMF werden in ein generisch einsetzbares Werkzeug zur Umschlüsselung von Pseudonymen und als Grundbaustein einer informationellen Gewaltenteilung eingearbeitet. Um eine robuste und praktikabel nutzbare Software zu erhalten, werden spätere Betreiber eng in die Entwicklung eingebunden.
- ▶ Die Mainzliste als Nachfolgeprodukt des PID-Generators wird dahingehend aktualisiert und erweitert, dass alle für eine Umsetzung des Datenschutzleitfadens der TMF notwendigen Funktionen unterstützt werden.
- ▶ Weiterentwicklung der Konzepte und Umsetzungen, insbesondere
  - ▶ Verbesserung anhand der Rückmeldungen von Anwenderworkshops,
  - ▶ Internationalisierung,
  - ▶ Einbindung oder Anbindung eines Kontaktmanagements (d. h. Funktionen zur Kontaktierung von Patienten über einen Treuhänder),
  - ▶ Anbindung eines Einwilligungsmanagements,

---

13 Mainzliste, Sampl.y.Auth und der Generische Informed Consent Service als Open-Source-Werkzeuge für Identitäts-, Einwilligungs- und Rechtsmanagement in der medizinischen Verbundforschung, vgl. <http://gepris.dfg.de/gepris/projekt/315057496>.

- ▶ Anbindung eines Authentifizierungsdienstes zur sicheren und komfortablen Nutzeranmeldung auch über Standorte hinweg sowie
- ▶ Interoperabilität zwischen den exemplarischen Werkzeugen in den Bereichen Pseudonymisierung, Kontaktmanagement, Einwilligungsmanagement, Authentifizierung und Web-basierte Registerwerkzeuge.

Über diese Fortentwicklungen wird in den TMF-Arbeitsgruppen IT-QM und Datenschutz regelmäßig berichtet, um rechtzeitig Rückmeldungen zukünftiger Nutzer zu berücksichtigen. Zusätzlich werden Architekten und Betreiber von Forschungsnetzen in einem öffentlichen Workshop die Möglichkeit zur Einflussnahme auf die Fortentwicklung der Werkzeuge erhalten.

Darüber hinaus besteht folgender offener Handlungsbedarf:

- ▶ Entwicklung und Anbindung sicherer Anonymisierungsfunktionen (z. B. zur k-Anonymisierung), siehe das Kapitel „Anonymisierung von medizinischen Individualdaten“ in diesem Report.
- ▶ Die TMF-Aktivitäten zur Schaffung einer angemessenen Aufmerksamkeit (Awareness) und zur Vermittlung des nötigen Know-hows (z. B. TMF-School, Workshops, Jahreskongresse, Tutorien auf Tagungen einschlägiger Fachgesellschaften) sind dauerhaft fortzuführen.
- ▶ Professioneller Betrieb als kommerzielle, möglichst preiswerte Dienstleistung alternativ zum Eigenbetrieb. Auf jeden Fall sollten Werkzeuge und Betrieb streng standardisiert sein. Die Möglichkeit einer formalen Zertifizierung durch geeignete Stellen (Behörden und Verbundforschung selbst) sollte genutzt werden.
- ▶ Problematisch für Dienstleistungsmodelle ist, dass diese Dienste nicht für alle Forschungsverbünde beim selben Dienstleister angesiedelt sein dürfen, weil ein solches „Superidentitätsmanagement“ nicht auf datenschutzrechtliche Akzeptanz hoffen könnte. Daher sollten mehrere unabhängige Dienstleister existieren. In der Erprobung ist ein Modell, bei dem verschiedene Forschungsverbünde gegenseitig Dienstleistungen anbieten. Aus dieser Problematik heraus ergibt sich ein dringender Standardisierungsbedarf, damit einzelne Standorte nicht verschiedene Schnittstellen für die Anbindung mehrerer Treuhänderservices umsetzen müssen.
- ▶ Das Angebot an fertigen Problemlösungen mit Service und Beratung ist auszubauen. Damit wird die Rechts- und Investitionssicherheit der medizinischen Forschungsverbünde nachhaltig gefördert.
- ▶ Ebenfalls wünschenswert ist eine systematische Evaluation der bereits eingesetzten Werkzeuge (Mainzliste, E-PIX, gPAS).

- ▶ Die Veröffentlichung und Pflege dieser Werkzeuge als Open-Source-Software erlaubt es prinzipiell jeder Forschergruppe, auch etwaige spezielle Anforderungen darin zu ergänzen. Damit ist eine praktikable Alternative zu Eigenentwicklungen geschaffen.
- ▶ Langfristig sollten die Schnittstellenkonzepte, die im deutschsprachigen Raum für das Identitätsmanagement in medizinischen Forschungsverbänden entwickelt wurden, auch in internationale Standards eingehen. Nur so wird perspektivisch eine einfache und sichere Integration in heterogene IT-Infrastrukturen, gerade auch im Bereich der Versorgung, möglich. Eine Herausforderung besteht allerdings darin, dass die hierzulande geforderte strikte Datentrennung zwischen Identitätsdaten auf der einen Seite und medizinischen Daten auf der anderen Seite bisher z. B. von den eher Dokumenten-orientierten IHE-Profilen in der „Quality, Research and Public Health Domain“ noch nicht vollständig unterstützt wird [12, S. 35f].

## Literatur

1. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
2. 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. Aufl. 2006, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
3. Becker, R., Ihle, P., Pommerening, K., Harnischmacher, U. *Ein generisches Datenschutzkonzept für Biomaterialbanken (Version 1.0)*. 2006. TMF, <http://www.tmf-ev.de/produkte/P010021> (Abruf: 2016-11-27).
4. Borg, A., Lablans, M., Ückert, F., *Mainzliste.client - Eine Bibliothek für den Zugriff auf Patientenlisten*, in *Tagungsband der 60. GMDS-Jahrestagung*. 2015.
5. Metschke, R., Wellbrock, R. *Datenschutz in Wissenschaft und Forschung*. 2002. Berliner Beauftragter für Datenschutz und Informationsfreiheit, <http://www.datenschutz-berlin.de/attachments/47/Materialien28.pdf> (Abruf: 2016-11-27).
6. DSK *Technische und organisatorische Anforderungen an die Trennung von automatisierten Verfahren bei der Benutzung einer gemeinsamen IT-Infrastruktur: Orientierungshilfe Mandantenfähigkeit*. 2012. Arbeitskreis Technische und organisatorische Datenschutzfragen der Konferenz der Datenschutzbeauftragten des Bundes und der Länder, <http://www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/2013/04/Mandantenf%C3%A4higkeit.pdf> (Abruf: 2016-11-27).
7. Lablans, M., Borg, A., Ückert, F., *A RESTful interface to pseudonymization services in modern web applications*. BMC Med Inform Decis Mak, 2015. **15**: S. 2.

8. Bialke, M., Bahls, T., Havemann, C., Piegsa, J., Weitmann, K., Wegner, T., Hoffmann, W., *MOSAIC – A Modular Approach to Data Management in Epidemiological Studies*. *Methods of Information in Medicine*, 2015. **54**(4): S. 364-371.
9. Bialke, M., Penndorf, P., Wegner, T., Bahls, T., Havemann, C., Piegsa, J., Hoffmann, W., *A workflow-driven approach to integrate generic software modules in a Trusted Third Party*. *J Transl Med*, 2015. **13**: S. 176.
10. Michael, J., *Doppelgänger gesucht. Ein Programm für kontextsensitive phonetische Textumwandlung*. *c't*, 1999. **25**: S. 252-261.
11. Postel, H.J., *Die Kölner Phonetik. Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse*. *IBM-Nachrichten*, 1969. **19**: S. 925-931.
12. IHE *Using IHE profiles for Healthcare - Secondary Data Access. IHE Quality, Research and Public Health (QRPH) White Paper*. 2016. Integrating the Healthcare Enterprise, Draft for Public Comment. Revision 1.0, [https://www.ihe.net/uploadedFiles/Documents/QRPH/IHE\\_QRPH\\_WP\\_Healthcare\\_Secondary\\_Data\\_Access\\_Rev1.0\\_PC\\_2016-07-25.pdf](https://www.ihe.net/uploadedFiles/Documents/QRPH/IHE_QRPH_WP_Healthcare_Secondary_Data_Access_Rev1.0_PC_2016-07-25.pdf) (Abruf: 2016-11-27).



# Anonymisierung von medizinischen Individualdaten

*Fabian Prasser, Murat Sariyar*

## Anwendungsbereich

Dieses Kapitel behandelt Werkzeuge für die Anonymisierung von medizinischen Individualdaten sowie Lösungen für die anonyme Analyse von Patienten- und Probandendaten in der medizinischen Forschung. Die Anonymisierung stellt neben der Pseudonymisierung eine weitere wesentliche Methode zum Schutz von Daten dar. Das Ziel von Anonymisierungsmethoden besteht darin, durch eine Veränderung oder Geheimhaltung personenbezogener Merkmale die Identität der Datensubjekte vor den Nutzern des Datensatzes zu verbergen und somit Missbrauch zu verhindern. Pseudonymisierte Daten, also Daten bei denen direkt identifizierende Merkmale von medizinischen Forschungsdaten getrennt sind, zeichnen sich dadurch aus, dass eine Reidentifikation der Forschungsdaten weiterhin möglich ist [1]. Die wesentliche Eigenschaft anonymisierter Daten besteht hingegen darin, dass ein solcher Rückbezug zum Patienten oder Probanden überhaupt nicht oder nur mit unverhältnismäßig hohem Aufwand möglich ist [2].

Mit Methoden der Anonymisierung ist es möglich, medizinische Forschungsdaten außerhalb der Bestimmungen des Datenschutzrechts zu verarbeiten, denn im rechtlichen Sinne handelt es sich bei anonymisierten Daten nicht mehr um personenbezogene Daten. Dies unterscheidet die Anonymisierung wesentlich von der Pseudonymisierung. Daher ist Anonymisierung vor allem für das „Data Sharing“ und die Sekundärnutzung von Daten relevant. Wesentliche Beispiele sind der einrichtungsübergreifende Austausch von Daten für Forschungszwecke und die Nutzung von im Behandlungskontext gesammelten klinischen Daten für die Forschung.

Darüber hinaus bestehen potentielle Synergien zwischen der Anonymisierung und der Pseudonymisierung von Daten. So können beispielsweise Methoden für die Berechnung von Reidentifikationsrisiken, die typischerweise zur Anonymisierung von Daten eingesetzt werden, dazu verwendet werden, zu bewerten, welcher Teil eines pseudonymisierten Datensatzes in welcher Form ohne größere Risiken verwendet werden kann. Auch können Forschern, die nur die Berechtigung zum Zugriff auf medizinische Daten besitzen, zusätzlich anonymisierte Teilmengen der Identitätsdaten zur Verfügung gestellt werden. Das größte Potential von Anonymisierungsverfahren liegt in diesem Kontext jedoch im weiterführenden Schutz medizinischer Daten, da auch als „medizinisch“

verstandene Merkmale von einem Angreifer mit entsprechendem Hintergrundwissen zur Reidentifikation genutzt werden können [3].

Fazit: Anonymisierungsmethoden für Patienten- oder Probandendaten in der medizinischen Forschung haben das Potential neue Anwendungsfälle zu erschließen und etablierte Schutzmaßnahmen, insbesondere die Pseudonymisierung, zu komplementieren und zu verbessern.

## Methoden und Anwender

Der Bereich der Datenanonymisierung umfasst ein breites Spektrum an Methoden, deren Ziel es ist, Daten vor einer Vielzahl von Angriffen zu schützen. Die bekannteste Bedrohung ist die Reidentifikation, also die Zuordnung von Datensätzen zu Datensubjekten. Darüber hinaus ist es unter Umständen aber auch ohne Reidentifikation möglich festzustellen, ob Daten über ein Subjekt in einem Datenbestand gespeichert sind oder abzuleiten, dass ein Individuum wahrscheinlich gewisse Eigenschaften hat (probabilistische Inferenz). In der Praxis ist es jedoch oftmals schwierig, solche über eine einfache Reidentifikation hinausgehenden Risiken stark zu reduzieren [4].

Ein erster Schritt bei der Anonymisierung von Daten kann durch Maskierung erfolgen. Unter diesem Sammelbegriff fasst man einfache Methoden zusammen, bei denen Identifikatoren, wie beispielsweise Namen, entfernt, verändert oder ersetzt werden. Maskierungsmethoden sind insbesondere auch bei der Entfernung von Identifikatoren aus unstrukturierten Texten relevant, wobei dort die Herausforderung in der Erkennung der zu maskierenden Informationen und nicht in der eigentlichen Maskierung selbst liegt. Im weiteren Verlauf dieses Kapitels werden ausschließlich Methoden zur Anonymisierung strukturierter Daten betrachtet.

Nationale und internationale Statistikinstitute gehören zu den Vorreitern beim Einsatz von Anonymisierungsverfahren [5]. Ein typischer Anwendungsfall ist hierbei die Anonymisierung von Daten im Rahmen der Erstellung von Bevölkerungsstatistiken. Auf methodischer Ebene kommen dabei zum einen Modelle zum Einsatz mit denen das Reidentifikationsrisiko eines Datensatzes gemessen werden kann und zum anderen Transformationsverfahren, mit denen Risiken durch die Veränderung von Merkmalsausprägungen reduziert werden können. Viele Verfahren aus diesem Bereich erhalten zwar wesentliche statistische Eigenschaften eines Datensatzes, sie sind aber nicht wahrheitserhaltend, d. h. sie verfälschen die Daten, beispielsweise durch das Hinzufügen von Rauschen [6]. Reidentifikationsrisiken werden meist nur soweit begrenzt, dass eine hohe Qualität der Ausgabedaten gewährleistet werden kann.

Motiviert durch eine Reihe erfolgreicher Reidentifikationsangriffe auf unzureichend anonymisierte Datenbestände in den USA (der wohl spektakulärste Fall war dabei die Reidentifikation der Patientenakte eines amerikanischen Politikers [7]) wurden von Forschern aus dem Bereich der Informatik stärkere Anonymisierungsverfahren entwickelt. Diese umfassen im Wesentlichen:

1. Methoden für die wahrheitsgetreue Veränderung von Merkmalsausprägungen, insbesondere durch Generalisierung. Bei diesem Verfahren werden beispielsweise Altersangaben durch Altersgruppen ersetzt.
2. Mathematische Modelle mit denen Datenschutzerfordernungen formal definiert werden können. Das bekannteste solche Modell ist die  $k$ -Anonymität [8]. Diese fordert, dass jeder Datensatz in einem Datenbestand bezüglich der Werte von Attributen mit denen ein hohes Reidentifikationsrisiko einhergeht von  $k-1$  anderen Datensätzen nicht unterschieden werden kann. Folglich ist es einem Angreifer meist nicht mehr möglich, einen Datensatz eindeutig einer Person zuzuordnen.

Anonymisierungsverfahren mit strikten Datenschutzerfordernungen basieren häufig auf einer automatischen Verarbeitung der Daten. Die im Bereich der Statistik entwickelten Methoden werden hingegen meist in einem iterativen, weitgehend manuellen, Prozess angewandt. Es wurden aber auch in der Statistik automatisierbare Verfahren für große Datenbestände entwickelt. Darüber hinaus können statistische Verfahren auch zur Erstellung statischer Maßnahmenkataloge für die Datenmaskierung eingesetzt werden. Ein wichtiges Beispiel hierfür ist die Safe Harbor Methode der HIPAA Privacy Rule aus den USA [9].

Neben Ansätzen bei denen ein Eingabedatensatz in einen anonymisierten Ausgabedatensatz transformiert wird, gibt es auch interaktive Verfahren zur anonymen Analyse von Daten. Bei diesem ebenfalls ursprünglich für statistische Datenbanken entwickelten Ansatz werden Daten sicher gespeichert und Anonymisierungsmethoden auf die Ergebnisse von Datenbankanfragen oder Auswertungen angewandt. Diese Herangehensweise unterscheidet sich wesentlich von klassischen Anonymisierungsverfahren, wie der  $k$ -Anonymisierung. Es können, beispielsweise durch Einsatz von Differential Privacy [10], deutlich höhere Schutzniveaus erreicht werden [11]. Ein Vorteil des Ansatzes besteht darin, dass keine Transformation der Daten notwendig ist und somit im Prinzip beliebige Analysen unterstützt werden können. Da der Zugriff auf die Daten aber stark eingeschränkt ist, was explorative Analysen der Daten zumindest erschwert, bestehen in der Praxis Akzeptanzprobleme.

Basierend auf Erfahrungen mit dem Einsatz von Anonymisierungsmethoden in der medizinischen Forschung insbesondere in den USA und in Kanada, sind einige wenige Methoden nach Einschätzung der Autoren von besonderer praktischer Relevanz. Dies sind 1) wahrheitsgetreue Transformationsmethoden, insbesondere die Generalisierung und Unterdrückung von Attributwerten, 2) mathematische Modelle zur Definition von Schutzniveaus, beispielsweise k-Anonymität, und 3) statistische Modelle zur Risikoabschätzung. Methoden, die über diese grundlegenden Ansätze hinausgehen, besitzen oft Eigenschaften, die ihren Einsatz in medizinischen Forschungsprozessen erschweren. So ist beispielsweise eine nicht-wahrheitsgetreue Transformation von Daten in der medizinischen Domäne nur schwer denkbar [11]. Maßnahmen gegen probabilistische Inferenz führen häufig zu einer zu starken Reduktion der Datenqualität [4, 12].

Dieser allgemeine Überblick über Anonymisierungsmethoden macht deutlich, dass diese an der Schnittstelle zwischen Informatik und Statistik angesiedelt sind und der fachgerechte Einsatz fundierte Kenntnisse und die Berücksichtigung des konkreten Kontextes erfordert. Die Anpassung von Methoden und Parametern an zu erwartende Angriffe und die Optimierung der Nützlichkeit der sich ergebenden Daten für spezifische Forschungsfragen sind dabei zentrale Aufgaben. Anwender von Anonymisierungsmethoden in der medizinischen Forschung benötigen fundiertes Fachwissen in den Bereichen Datenschutz, Datenmanagement und Datenanalyse, um diese Abwägungen zwischen Reidentifikationsrisiko und Nützlichkeit vornehmen zu können.

Professionelle Beratungsangebote und Dienstleister im Bereich Datenanonymisierung sind selten oder wie die Data Analysis OG aus Österreich<sup>1</sup> nicht auf die medizinische Domäne spezialisiert. Die TMF hat sich daher vorgenommen, Schulungen durchzuführen, in denen Teilnehmer praktische Erfahrungen mit Anonymisierungsmethoden und –werkzeugen sammeln können. In 2016 wurden drei Schulungen durchgeführt, welche die Autoren dieses Kapitels mit ausgearbeitet haben. Da die Veranstaltungen in der Community auf großes Interesse gestoßen sind, ist es geplant, ein regelmäßiges Schulungsangebot zu etablieren.

Fazit: Die Anonymisierung von Daten erfordert fundiertes Fachwissen sowie Erfahrung im Datenmanagement und der Datenanalyse. Beratungsangebote und Dienstleister sind selten. Die TMF bietet entsprechende Schulungen an, die weiter ausgebaut werden sollen.

---

1 <http://www.data-analysis.at/>

## Rahmenbedingungen

Die rechtlichen Rahmenbedingungen für die Datenanonymisierung sind vor allem im Datenschutzrecht festgelegt. Zum einen beschreibt die Anonymisierung eine Möglichkeit dem Gebot der Datensparsamkeit zu genügen, indem Daten anonymisiert werden, sobald dies im Rahmen eines Forschungsvorhabens vertretbar ist. Zum anderen ist es mit Methoden der Anonymisierung möglich, Daten außerhalb der Bestimmungen des Datenschutzrechts zu verarbeiten. Dies liegt darin begründet, dass es sich im rechtlichen Sinne bei anonymisierten Daten nicht mehr um personenbezogene Daten handelt. Dies ermöglicht eine deutlich größere Freiheit bei der Datenverarbeitung und erschließt neue Anwendungsfälle. Problematisch ist in diesem Zusammenhang die Tatsache, dass die Gesetzgebung einen großen Interpretationsspielraum zulässt. So ist es beispielsweise nicht klar, ob Daten nur vor Reidentifikation geschützt werden müssen, oder ob auch Maßnahmen gegen probabilistische Inferenz notwendig sind. Dies ist ein wesentlicher Hemmfaktor für die Etablierung von Anonymisierungsmethoden in der medizinischen Forschung.

Für die rechtliche Bewertung von konkreten Anonymisierungsmethoden und deren Einsatz sind zwei wesentliche Eigenschaften relevant. Zum einen muss zwischen absoluter und de-facto Anonymität unterschieden werden. Absolute Anonymität ist gegeben, wenn unter keinen Umständen eine Reidentifizierung möglich ist. De-facto Anonymität, die auch im BDSG definiert ist, berücksichtigt den Kontext der Datennutzung und verlangt, dass die vorhandenen Daten „nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbar natürlichen Person zugeordnet werden können“.<sup>2</sup> Ein weiterer Aspekt des Kontextbezugs ist, wem gegenüber Daten anonym sein müssen. So spricht man von „relativer Anonymisierung“ bzw. „relativem Personenbezug“, wenn lediglich verlangt wird, dass nur externe Datennutzer keinen Bezug herstellen können, dies aber beispielsweise dem ursprünglichen Datenhalter möglich ist. Der Europäische Gerichtshof (EuGH) hat in einem aktuellen Urteil zur Personenbeziehbarkeit von IP-Adressen grundsätzlich eine relative Sicht der Personenbeziehbarkeit unterstützt, allerdings sehr hohe Hürden für eine Umsetzung aufgestellt [13]. Ob und wie dieses Urteil die Rechtspraxis in Deutschland beeinflusst, bleibt jedoch abzuwarten.

Jenseits der kategorialen Frage eines geltenden Personenbezugs und der daraus folgenden Anwendbarkeit des Datenschutzrechts können Anonymisierungsmethoden auch innerhalb des Anwendungsbereichs der Datenschutzregularien sinnvoll einge-

---

2 Siehe [https://www.gesetze-im-internet.de/bdsg\\_1990/\\_3.html](https://www.gesetze-im-internet.de/bdsg_1990/_3.html).

setzt werden. Dies ist insbesondere im Rahmen der sogenannten Forschungsklauseln möglich, die eine Abwägung von allgemeinen Forschungsinteressen einerseits und den zu schützenden Persönlichkeitsrechten der Betroffenen andererseits im Einzelfall vorsehen. Hier kann es Fälle geben, in denen das Forschungsinteresse das Interesse der Individuen am Ausschluss ihrer Daten von der Verarbeitung so erheblich überwiegt, so dass eine Verarbeitung und Nutzung der Daten auch nach Datenschutzrecht und ohne Einwilligung rechtmäßig sein kann (vgl. § 28 Abs. 6 Nr. 4 BDSG). Der in diesen Fällen notwendige Abwägungsprozess kann nun durch die hier dargestellten Anonymisierungsmethoden stark im Sinne einer Absenkung der Individualinteressen der Betroffenen beeinflusst werden, auch wenn keine vollständige Anonymisierung im gesetzlichen Sinne erreicht wird.

**Im Gegensatz zur Datenpseudonymisierung existieren für die Anonymisierung keine Richtlinien oder Leitfäden. Dies macht einen Einsatz schwierig, insbesondere aufgrund fehlender Rechtssicherheit. Ein Beispiel dafür, wie ein solches Regelwerk aussehen könnte, ist die HIPAA Privacy Rule aus den USA [9].** Neben der Verwendung der bereits erwähnten einfachen Regeln für die Datenmaskierung erlaubt das Gesetz alternativ auch den Einsatz von Anonymisierungsverfahren.

**Die technisch-methodischen Rahmenbedingungen im Bereich der Datenanonymisierung können als ausreichend bewertet werden. Für die Unterstützung von automatisierten sowie manuellen Anonymisierungsprozessen und Risikoanalysen ist eine Vielzahl von Werkzeugen verfügbar. Darüberhinaus wurden spezifische Lösungen für einzelne Anwendungsfälle der medizinischen Forschung entwickelt (bspw. [14]).**

Fazit: Es existieren keine Leitfäden für die Anonymisierung medizinischer Daten und bedingt durch den großen Interpretationsspielraum der Gesetzgebung und einigen offenen Fragen besteht beim Einsatz von Anonymisierungsmethoden keine ausreichende Rechtssicherheit. Es existiert jedoch eine Vielzahl relevanter Methoden und Implementierungen mit denen Daten sinnvoll vor realistischen Bedrohungen geschützt werden können.

## Kategorisierung der IT-Landschaft

Die IT-Landschaft im Bereich der Anonymisierungswerkzeuge besteht vor allem aus Open-Source-Software. Lösungen können dahingehend unterschieden werden, ob sie eine anonymisierte Version eines Eingabedatensatzes für die weitergehende Analyse erzeugen oder die Analyse eines Datensatzes in anonymer Form ermöglichen.

Lösungen der ersten Kategorie sind insbesondere im Bereich der amtlichen Statistik entwickelt worden und zeichnen sich häufig durch einen sehr interaktiven Anonymisierungsprozess aus. Sie implementieren sowohl wahrheitserhaltende als auch nicht-wahrheitserhaltende Transformationsmethoden und stellen Methoden zur Modellierung von Haushalts- und Familienstrukturen in Datensätzen zur Verfügung. Auch bieten viele Werkzeuge eine Anbindung an Statistiksoftware, wie beispielsweise R. Es gibt aber auch Lösungen, die einen weitgehend automatisierten Anonymisierungsprozess ermöglichen. Diese konzentrieren sich auf wahrheitserhaltende Transformationsmethoden und unterstützen ein breites Spektrum an Ansätzen für die Analyse der Ergebnisqualität. Neben wenigen ausgereiften Lösungen gibt es auch eine große Anzahl von Forschungsprototypen, die sich nicht für einen Praxiseinsatz eignen.

Lösungen für die Analyse von Daten in anonymer Form lassen im Regelfall nur einen direkten Zugriff auf die Ergebnisse von Auswertungen, nicht aber auf die eigentlichen Individualdaten selbst zu. Da auch hochaggregierte Analyseergebnisse für eine Reidentifikation von Patienten oder Probanden verwendet werden können [15], werden sowohl wahrheitserhaltende als auch nicht-wahrheitserhaltende Transformationsmethoden verwendet, um Ergebnisse von Analysen oder Datenbankabfragen zu anonymisieren.

### Aktueller Bestand<sup>3</sup>

Die kanadische Firma Privacy Analytics<sup>4</sup> vertreibt das einzige kommerziell verfügbare Anonymisierungstool unter dem Namen CORE. Es zählt zu den wenigen Werkzeugen die explizit für die medizinische Domäne entwickelt wurden. Es implementiert k-Anonymität mit Generalisierung und Unterdrückung von Attributwerten [16]. Darüber hinaus stellt es Methoden für die Analyse von Reidentifikationsrisiken und die risikobasierte Parametrisierung von Anonymisierungsverfahren zur Verfügung. CORE wurde für den kanadischen und den US-Markt entwickelt und ist für die dort gültige Datenschutzgesetzgebung, insbesondere die HIPAA Privacy Rule, optimiert. Ein Alleinstellungsmerkmal der Software ist, dass sie ähnlich wie gängige ETL-Tools die Möglichkeit bietet, Workflows zu definieren, mit denen Daten aus unterschiedlichen Datenbanken und Datenbanktabellen extrahiert, kombiniert und anonymisiert werden können.

ARX<sup>5</sup> ist eine Open-Source-Lösung die von einem der Autoren dieses Kapitels entwickelt wird und die auch im Rahmen der Schulungen der TMF eingesetzt wurde. Im

---

3 Einige der hier dargestellten Lösungen wurden auch im Rahmen eines TMF-Workshops im März 2015 vorgestellt. Ein Nachbericht mit den gezeigten Folien ist unter [www.tmf-ev.de/news/1706](http://www.tmf-ev.de/news/1706) verfügbar.

4 Siehe <http://www.privacy-analytics.com/software/privacy-analytics-core/>.

5 Siehe <http://arx.deidentifier.org>.

Vergleich mit anderen Anonymisierungswerkzeugen zeichnet es sich durch eine umfangreiche graphische Benutzeroberfläche, eine hohe Skalierbarkeit und eine breite Methodenunterstützung aus. ARX implementiert ähnliche Anonymisierungsmethoden wie die Software von Privacy Analytics, unterstützt darüber hinaus aber weitere Datenschutzmodelle, beispielsweise zum Schutz vor Inferenz, und ein größeres Spektrum an Transformationsmethoden [17]. Ein weiteres Alleinstellungsmerkmal von ARX ist, dass es in der Lage ist, auch Datensätze mit einer großen Anzahl an risikobehafteten Attributen automatisiert zu verarbeiten. Die Software kann zwar Daten aus Tabellenkalkulationsprogrammen und relationalen Datenbanksystemen importieren, unterstützt aber im Gegensatz zu CORE keine ETL Prozesse.

sdcmicro<sup>6</sup> ist ein Paket für die Statistiksoftware R. Ähnlich wie die Software von Privacy Analytics fokussiert es auf Maßnahmen zum Schutz vor Reidentifikation [18], es unterstützt aber deutlich mehr Transformationsmethoden. Durch die Integration in R steht ein breites Spektrum an Funktionen für das Importieren und Exportieren von Daten sowie für die Analyse von Ein- und Ausgabedaten zur Verfügung. Der Anonymisierungsprozess muss schrittweise manuell durchgeführt werden, bevorzugt über die Konsole des R-Systems. sdcmicro stellt zwar eine graphische Benutzeroberfläche zur Verfügung, deren Verwendung die Skalierbarkeit der Software aber deutlich reduziert. Die Stärken des Werkzeugs umfassen insbesondere eine hohe Skalierbarkeit (falls statt der graphischen Benutzeroberfläche, die konsolenbasierte Version genutzt wird), sowie eine umfangreiche Unterstützung von Risikomodellen und nicht-wahrheitserhaltenden Transformationsoperationen. Ein Alleinstellungsmerkmal ist, dass aufgrund der direkten Integration in eine entsprechende Software eine Vielzahl von Methoden zur begleitenden Analyse der statistischen Validität und Nützlichkeit der Daten verfügbar ist.

$\mu$ -Argus<sup>7</sup> ist eine weitere quelloffene Software, die von Statistikern entwickelt wurde. Ähnlich wie sdcmicro implementiert sie einen sehr flexiblen Anonymisierungsprozess, der viel manuelle Interaktion benötigt. Die Software besitzt eine ausgereifte graphische Benutzeroberfläche.  $\mu$ -Argus implementiert als einzige Anwendung im Feld keine Datenschutzmodelle, wie  $k$ -Anonymität. Auch bietet es nur wenige Möglichkeiten zur Evaluation der Ergebnisqualität. Weitere Forschungsprototypen und verfügbare Implementierungen von Methoden der Datenanonymisierung umfassen: die UTD Anonymization Toolbox<sup>8</sup>, das Cornell Anonymization Toolkit<sup>9</sup> und das TMF Anon-

---

6 Siehe <https://cran.r-project.org/web/packages/sdcmicro/index.html>.

7 Siehe <http://neon.vb.cbs.nl/casc/mu.htm>.

8 Siehe <http://www.cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php>.

9 Siehe <http://sourceforge.net/projects/anony-toolkit/>.

Tool<sup>10</sup>. Aufgrund ihrer eingeschränkten Funktionalität, geringer Interoperabilität und der prototypischen Natur der Implementierung eignet sich keines dieser Werkzeuge für einen Einsatz in der Praxis.

Als Alternative zur Anonymisierung von Primärdaten gibt es Werkzeuge, die eine interaktive Analyse von Daten in anonymer Form ermöglichen. Die Firma Aircloak<sup>11</sup> vertreibt in diesem Bereich eine kommerzielle Lösung. Die Ergebnisse von Datenbankabfragen werden durch verschiedene Verfahren anonymisiert. Eines dieser Verfahren ist das Hinzufügen von Rauschen, was verwandt ist mit typischen Methoden zur Implementierung von Differential Privacy [10]. Zusätzlich werden an das System gestellte Anfragen überwacht, um eine Inferenz von personenbezogenen Daten durch die Kombination verschiedener Anfrageergebnisse zu verhindern. Die Open-Source-Lösung DataSHIELD<sup>12</sup> sowie die Software Sharemind<sup>13</sup> der Firma Cybernetica funktionieren beide ähnlich. Sie sind aber vor allem für eine integrierte Analyse verteilter Datenbestände konzipiert worden. Beide Systeme stellen dem Nutzer eine modifizierte R-Umgebung zur Verfügung. Sharemind verwendet kryptographische Methoden zur sicheren Integration verteilter Datenbestände, wohingegen DataSHIELD ausschließlich die Ergebnisse von parallel durchgeführten Einzelanalysen integriert. Zur Anonymisierung der Analyseergebnisse wird bei Sharemind, ähnlich wie bei der Lösung der Firma Aircloak, Rauschen hinzugefügt. Im Gegensatz dazu setzt DataSHIELD auf klassische Anonymisierungsverfahren für multidimensionale Kontingenztabellen, wie sie auch von entsprechenden Varianten der Lösungen sdcMicro und  $\mu$ -Argus, genannt sdcTable<sup>14</sup> und  $\tau$ -Argus<sup>15</sup>, implementiert werden.

## Bewertung und Handlungsbedarf

Anonymisierungsmethoden haben das Potential, neue Anwendungsfälle zu erschließen, etablierte Schutzmaßnahmen zu komplementieren und datenintensive Forschungsprozesse sicherer zu machen. Die allgemeine Situation hinsichtlich der Einsetzbarkeit entsprechender Methoden in der Praxis kann derzeit folgendermaßen bewertet werden:

- Erfahrungen aus anderen Ländern zeigen, dass der Einsatz von Methoden zum Schutz vor Reidentifikation in der Praxis sinnvoll ist [19].

---

10 Siehe [http://www.tmf-ev.de/Themen/Projekte/V08601\\_AnonTool.aspx](http://www.tmf-ev.de/Themen/Projekte/V08601_AnonTool.aspx).

11 Siehe <https://www.aircloak.com/downloads/aircloak-white-paper.pdf>.

12 Siehe <http://www.datashield.ac.uk/>.

13 Siehe <https://sharemind.cyber.ee/>.

14 Siehe <https://cran.r-project.org/web/packages/sdcTable/index.html>.

15 Siehe <http://neon.vb.cbs.nl/casc/tau.htm>.

- ▶ **Es ist grundsätzlich möglich, Analyseverfahren aus der medizinischen Forschung mit anonymisierten Daten durchzuführen [14].**
- ▶ Praktische Erfahrungen sind in Deutschland aufgrund der geringen Verbreitung aber selten.
- ▶ Das Fehlen von Leitfäden und rechtlichen Vorgaben ist, ebenso wie die Tatsache, dass ein fundiertes Fachwissen zum korrekten Einsatz von Anonymisierungsmethoden unerlässlich ist, ein wesentlicher Hemmfaktor für einen breiteren Einsatz.
- ▶ Bekannte Schwächen gängiger Anonymisierungsverfahren oder gewisse Kontexte (siehe unten) machen es unter Umständen notwendig, zusätzliche Sicherheitsmaßnahmen zu ergreifen [20]:
  - ▶ Um Anonymisierungsmethoden für einen spezifischen Anwendungsfall anpassen zu können, kann es sinnvoll sein, den Zugriff auf einen vorher klar definierten Nutzerkreis zu beschränken.
  - ▶ Darüber hinaus können Data Use Agreements eingesetzt werden, die es dem Empfänger der Daten untersagen, eine Reidentifikation vorzunehmen.
  - ▶ Zusätzlich können Patienten oder Probanden im Rahmen der informierten Einwilligung auf die mit einer Datenanonymisierung einhergehenden Restriktionen hingewiesen werden.
- ▶ Bei hochdimensionalen (50 Attribute werden häufig als Grenze genannt) und gleichzeitig sensiblen Informationen, wie beispielsweise genomischen Daten, stoßen Anonymisierungsverfahren an ihre Grenzen. In diesem Fall kann der Einsatz von kryptographischen Methoden, beispielsweise von Protokollen für sichere Mehrparteienberechnungen, oder problemspezifischen Implementierungen von Differential Privacy erwogen werden [21].

Neben dieser allgemeinen Bewertung ist es sinnvoll, zwischen der Verwendung von Anonymisierungsmethoden im Rahmen einer Hypothesengenerierung und der Verwendung in Forschungskontexten mit strengen Anforderungen an die Validität von Analyseergebnissen zu unterscheiden.

Im ersten Fall, also der Bereitstellung von großen Datenmengen für die Mustererkennung oder Hypothesengenerierung, welcher zumeist mit Big-Data-Szenarien assoziiert wird, bestehen zwar zum einen sehr hohe Anforderungen an den Datenschutz, zum anderen aber weniger große Anforderungen an die analytische Validität der Daten. Darüber hinaus können bei hohen Fallzahlen mit gängigen Anonymisierungsverfahren bessere Ergebnisse hinsichtlich der Datenqualität erzielt werden [14]. Deshalb ist gerade in diesem Bereich die Verwendung von Anonymisierungsmethoden angebracht und sollte durch Best-Practices beschrieben werden.

Im zweiten Fall sollte die Anonymisierung in das Forschungsdatenmanagement<sup>16</sup> so eingebettet werden, dass eine optimale Abwägung zwischen den notwendigen Schutzmaßnahmen und den Anforderungen an die Qualität der Daten möglich ist. So kann beispielsweise ein geringes Schutzniveau akzeptabel sein, wenn Daten mit strikter Zugangsbeschränkungen für eine überwachte Analyse durch eine vertrauenswürdige Person zugänglich gemacht werden. Um solche notwendigen Kombinationen von primären und sekundären technischen Maßnahmen sowie organisatorischen Maßnahmen zu ermöglichen, sollte man typische Anwendungsszenarien zusammen mit notwendigen technischen und organisatorischen Schutzmechanismen in generischen Datenschutzkonzepten feingranular abbilden. Wichtig kann es zudem sein, dass der Einfluss von an den Daten vorgenommenen Veränderungen und den dadurch hervorgerufenen statistischen Verzerrungen bei Auswertungen berücksichtigt wird. Dies erfordert entsprechendes Fachwissen [22].

Da der datenschutzkonforme Austausch und die Integration von Daten, auch über die Grenzen einzelner Standorte hinweg, eine wesentliche Voraussetzung für innovative medizinische Forschungsprozesse darstellt [23] und Anonymisierung einen wichtigen Beitrag zur Reduktion von Risiken leisten kann, sehen wir folgenden Handlungsbedarf:

- ▶ Die offenen rechtlichen Fragen sind so weit zu klären, dass Rechtssicherheit bezüglich der Anwendung von Anonymisierungsverfahren erreicht wird.
- ▶ Es müssen Leitfäden erstellt und Fachwissen aufgebaut werden. Der TMF-Workshop zum Thema Datenanonymisierung im März 2015<sup>17</sup> und der Inhalt der TMF-Schulungen zur Datenanonymisierung im Jahr 2016<sup>18</sup> bilden gute Grundlagen dafür.
- ▶ Die Einsatzfähigkeit von Anonymisierungswerkzeugen in medizinischen Forschungsprozessen muss weiter evaluiert werden, um eine bessere Einschätzung existierender Lösungen zu erhalten und ggf. Verbesserungen vorzunehmen.

---

16 Siehe das Kapitel zum Forschungsdatenmanagement im vorliegenden Report.

17 Siehe [www.tmf-ev.de/news/1706](http://www.tmf-ev.de/news/1706).

18 Siehe <http://www.tmf-ev.de/Themen/Projekte/V08602MWSAnonymisierung.aspx>.

## Literatur

1. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
2. *ISO Health informatics - Pseudonymization*. 2008. International Organization for Standardization, ISO/TS 25237:2008.
3. Loukides, G., Denny, J.C., Malin, B., *The disclosure of diagnosis codes can breach research participants' privacy*. *J Am Med Inform Assoc*, 2010. **17**(3): S. 322-327.
4. El Emam, K., Álvarez, C., *A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques*. *International Data Privacy Law*, 2015. **5**(1): S. 73-87.
5. Dalenius, T., *Towards a methodology for statistical disclosure control*. *Stat Tidskr*, 1977. **15**: S. 429-444.
6. Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K., de Wolf, P.-P., *Statistical Disclosure Control*. 1. Aufl. 2012, Wiley.
7. Sweeney, L., *k-Anonymity: A model for protecting privacy*. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2002. **10**(05): S. 557-570.
8. Samarati, P., *Protecting Respondents' Identities in Microdata Release*. *IEEE Trans. Knowl. Data Eng.*, 2001. **01/2001**(13): S. 1010-1027.
9. *US Summary of the HIPAA Privacy Rule*. 2003. U.S. Department of Health & Human Services, <http://www.hhs.gov/sites/default/files/privacysummary.pdf> (Abruf: 2016-11-27).
10. Dwork, C., *Differential privacy*, in *Encyclopedia of Cryptography and Security*, Hrsg.: H.C.A. van Tilborg and S. Jajodia. 2011, Springer, New York. S. 338-340.
11. Dankar, F.K., El Emam, K., *Practicing differential privacy in health care: a review*. *Transactions on Data Privacy*, 2013. **5**: S. 35-67.
12. Brickell, J., Shmatikov, V. *The cost of privacy: destruction of data-mining utility in anonymized data publishing*. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2008. Las Vegas, Nevada, USA: S. 70-78.
13. *EuGH Urteil des Gerichtshofs (Zweite Kammer) vom 19. Oktober 2016. Patrick Breyer gegen Bundesrepublik Deutschland. Aktenzeichen C-582/14*. 2016. Europäischer Gerichtshof.
14. Heatherly, R.D., Loukides, G., Denny, J.C., Haines, J.L., Roden, D.M., Malin, B.A., *Enabling genomic-phenomic association discovery without sacrificing anonymity*. *PLoS One*, 2013. **8**(2): S. e53875.

15. Homer, N., Szelling, S., Redman, M. et al., *Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays*. PLoS Genet, 2008. **4**(8): S. e1000167.
16. El Emam, K., Dankar, F.K., Issa, R. et al., *A globally optimal k-anonymity method for the de-identification of health data*. J Am Med Inform Assoc, 2009. **16**(5): S. 670-682.
17. Prasser, F., Kohlmayer, F., Lautenschläger, R., Kuhn, K.A., *ARX – A Comprehensive Tool for Anonymizing Biomedical Data*. AMIA Annu Symp Proc, 2014. **2014**: S. 984-993.
18. Templ, M., Kowarik, A., Meindl, B., *Statistical Disclosure Control for Micro-Data Using the R Package sdcMicro*. Journal of Statistical Software, 2015. **67**(4): S. 1-36.
19. McGraw, D., *Building public trust in uses of Health Insurance Portability and Accountability Act de-identified data*. J Am Med Inform Assoc, 2013. **20**(1): S. 29-34.
20. Rubinstein, I.S., Hartzog, W. *Anonymization and Risk*. 2015. New York University School of Law: Public Law & Legal Theory Research Paper Series, Working Paper NO. 15-36, [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2646185](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2646185) (Abruf: 2016-11-27).
21. Naveed, M., Ayday, E., Clayton, E.W. et al., *Privacy in the Genomic Era*. ACM Comput Surv, 2015. **48**(1).
22. Shlomo, N., *Statistical Disclosure Limitation for Health Data: A Statistical Agency Perspective*, in *Medical Data Privacy Handbook*, Hrsg.: A. Gkoulalas-Divanis and G. Loukides. 2015, Springer, Heidelberg. S. 201-230.
23. Malin, B., Karp, D., Scheuermann, R.H., *Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research*. J Investig Med, 2010. **58**(1): S. 11-18.



# Cloud-Computing in der medizinischen Forschung

*Johannes Drepper*

## Anwendungsbereich

Die Nutzung von Cloud-Computing wird heute als wesentlicher IT-Trend angesehen [1, 2], der eine hohe Flexibilität und Kosteneffizienz ermöglicht und daher auch Auswirkungen auf IT-Infrastrukturen in der Forschung hat [3]. Dies gilt in besonderem Maße für die biomedizinische Forschung, die in den letzten Jahren von einem starken Anstieg der Anforderungen an Speicher- und Rechenkapazität gekennzeichnet ist [4]. Technisch gesehen ist die Cloud ein virtualisiertes Angebot von IT-Services. Dabei können Hardwareressourcen, im wesentlichen Rechenkapazität und Speicherplatz (Infrastructure as a Service – IaaS), Anwendungs- und Entwicklungsplattformen (Platform as a Service – PaaS) als auch fertige Anwendungen (Software as a Service – SaaS) über Netzwerkverbindungen angeboten werden. Diese Angebote bauen im Regelfall hierarchisch aufeinander auf: Für die Nutzung einer über das Netzwerk angebotenen Auswertungssoftware muss der Anbieter auch entsprechende Hardwareressourcen, auf denen die Anwendung ausgeführt wird, vorhalten.

Nach der einschlägigen Definition des US National Institute of Standards and Technology (NIST) zeichnen sich Cloud-basierte Infrastrukturen durch folgende Eigenschaften aus [5]:

1. Die Bereitstellung von Ressourcen erfolgt hoch automatisiert (on-demand, self service).
2. Ein netzwerkbasierter Zugang zu Ressourcen erfolgt über standardisierte Schnittstellen.
3. Die vorhandenen Ressourcen werden dynamisch auf die anfragenden Nutzer verteilt, unabhängig von der geografischen Position der Ressourcen und der Nutzer.
4. Ressourcen können kurzfristig hinzugebucht oder freigegeben werden (Elastizität).
5. Die Ressourcennutzung wird protokolliert und abgerechnet.

Ein häufig genanntes Anwendungsfeld für Cloud-Computing in der biomedizinischen Forschung ist der Bereich der Speicherung und Verarbeitung der sogenannten Omics-Daten. Für diesen Bereich kann die extreme Zunahme der Datenmengen in den letzten Jahren und der daraus resultierende Anstieg der notwendigen Speicher- und Rechen-

kapazität<sup>1</sup> als Treiber für den Trend hin zur Cloud angesehen werden [6-8]. Dabei ist jedoch auch nachvollziehbar, dass die Zunahme der Datenmengen auch zum gegenteiligen Trend führt, die Daten zunehmend dort zu lassen und zu verarbeiten, wo sie auch erhoben wurden. Entsprechend des Cloud-Paradigmas kann dann weltweit der Zugriff und die effiziente Verarbeitung und Auswertung der Daten angeboten werden [7]. In diesem Fall ist aber der physische Standort der Ressourcen ein kritischer Parameter und nicht mehr so flexibel, wie es das dritte NIST-Kriterium (s. o.) eigentlich nahelegt. Insbesondere können die Daten aufgrund ihrer Menge eben nicht mehr schnell auf Server an anderen Standorten verschoben werden. Noch zu wenig beachtet erscheint die Anforderung, die in diverse Auswertungen und Studien eingegangenen genetischen Daten auch langfristig, sicher und ausreichend gut dokumentiert für die wissenschaftlichen Nachvollziehbarkeit und ggf. auch für die Nachnutzbarkeit aufzubewahren [9]. Auch hier könnten zentralisierte und Cloud-basierte Lösungen zum Einsatz kommen und langfristig helfen, die Kosten in einem vertretbaren Rahmen zu halten.

Aber auch unabhängig von der Bereitstellung großer Mengen genetischer Daten in der Cloud profitiert die bioinformatische Entwicklung von der Möglichkeit, kurzfristig und kostengünstig größere Mengen an Computing-Ressourcen nutzen zu können, als dies lokal typischerweise möglich ist [6]. Dies betrifft nicht nur, aber auch und gerade die Entwicklung, Testung und Validierung neuer Verarbeitungs- und Auswertungs-algorithmen für genetische Daten [vgl. 10, 11, 12]. Hierfür werden im internationalen Umfeld unter anderem die Angebote großer und kommerzieller US-amerikanischer Anbieter genutzt [6], auch wenn dies nicht in jedem Fall mit den europäischen Datenschutznormen vereinbar erscheint (s. u.).

Auch jenseits der Speicherung und Verarbeitung von Omics-Daten kommen im medizinischen Umfeld Clouds zum Einsatz. Beispielhaft kann hier auf das Textmining großer Mengen unstrukturierter klinischer Daten in der Cloud zu Forschungs- und Qualitätssicherungszwecken [13] verwiesen werden. Den spezifischen Vorteilen und Risiken der Nutzung von Clouds für das Datenmanagement in klinischen Studien hat sich ein Workshop des European Clinical Research Infrastructures Network (ECRIN) gewidmet [14].<sup>2</sup> In einem Review aus 2015 kommen Griebel et al. allerdings zu dem Schluss, dass viele Lösungen, die als Cloud-basiert publiziert werden, bei näherem Blick den oben dargestellten NIST-Kriterien nicht vollständig entsprechen [15]. Insofern kann davon ausgegangen werden, dass die Bedeutung der Cloud-Technologie in einigen Bereichen überschätzt wird. Dies gilt zumindest dann, wenn man die

---

1 Siehe hierzu auch die ausführliche Darstellung der IT-Anforderungen in diesem Bereich im entsprechenden Kapitel in diesem Report.

2 Vgl. hierzu auch das Kapitel zu klinischen Studien im vorliegenden Report.

NIST-Kriterien als Maßstab nimmt und klassische IT-Services, die diesen Kriterien nicht entsprechen, außen vor lässt.

Wie im Weiteren noch gezeigt wird, ergibt sich jedoch eine mögliche Begrenzung der Anwendbarkeit des Cloud-Paradigmas auch aus den europäischen und deutschen Datenschutznormen, die gerade den Vorteil der flexiblen (vgl. NIST-Kriterium Nr. 1) und geografisch nicht vorhersagbaren Ressourcen-Nutzung (vgl. NIST-Kriterium Nr. 3) stark relativieren [vgl. 16].

Gleichwohl zeigt sich gerade für die vernetzte medizinische Forschung, die auch heute schon eine zentrale Zusammenführung von Daten und Auswertungsmöglichkeiten mit den hierfür erforderlichen Regelungen und Vertrauensstellungen erfordert, ein deutliches Potential für die Nutzung von Clouds. Eine spannende Frage ist dabei auch, ob die akademische und öffentlich geförderte Forschungswelt hauptsächlich auf der Nutzerseite agieren oder künftig auch im größeren Umfang Cloud-Services anbieten wird. Als weiterer relevanter Faktor für die künftige Entwicklung muss zudem die zunehmende Einbindung von Patienten und ihrer mobilen Devices berücksichtigt werden.<sup>3</sup> Vor dem Hintergrund der heute schon stark Cloud-orientierten Infrastrukturen in diesem Bereich kann auch dieser Trend zu einem Anstieg medizinischer Forschungsdaten in der Cloud führen.

## Betreiber, Anwender, Nutzer

Gemäß der oben bereits eingeführten Publikation des NIST [5] werden die folgenden Organisationsformen von Clouds unterschieden:

1. Private,
2. Community,
3. Public und
4. Hybrid Cloud.

Eine Private Cloud steht exklusiv nur einer Organisation zur Verfügung. Sie kann von dieser Organisation oder einem externen Anbieter betrieben werden. Bei der Community Cloud weitert sich der Nutzerkreis zu einer mehr oder minder organisierten Gemeinschaft von Organisationen und Einrichtungen mit gleichsinnigen Interessen in Bezug auf die Cloud-Nutzung. Eine Community Cloud kann von einer oder mehreren der nutzenden Organisationen betrieben werden oder auch von einem externen Anbie-

---

<sup>3</sup> Vgl. hierzu auch das Kapitel zu mobilen IT-Werkzeugen im vorliegenden Report.

ter. Die Public Cloud steht hingegen grundsätzlich jedem Benutzer zur Verfügung und kann von öffentlichen Einrichtungen oder auch privaten, kommerziellen Anbietern gehostet werden. Grundsätzlich sind auch Kombinationen dieser Modelle innerhalb einer Infrastruktur denkbar – in diesen Fällen wird von einer Hybrid Cloud gesprochen.

Bestimmte Vorteile des Cloud-Computings, wie die flexible und skalierbare Nutzung von Ressourcen wie Rechenleistung und Speicherplatz bis hin zu vollständigen Anwendungen kommen typischerweise insbesondere bei Public oder Community Clouds zum Tragen: Die Nutzer können hier von einer Kostenumverteilung von Investitions- zu Betriebsaufwänden profitieren. Dies führt im Idealfall aufgrund einer zielgenauen Mittelallokation zu einer Kostensenkung. Allgemeine Versprechungen von Kostensenkungen sind jedoch immer zu hinterfragen. Im Einzelfall hängt der Kostenvergleich von sehr vielen anwendungsfall-spezifischen Parametern ab [vgl. z. B. 17]. Wenn die für ein Cloud-Angebot nötige Hard- und Software selber vorgehalten werden muss, wie dies bei einer Private Cloud häufig der Fall ist, entfallen diese Vorteile. Andere Vorteile, wie z. B. standardisierte Systemumgebungen, eine flexiblere und zielgenauere Nutzung eigener Hardwareressourcen und ein vereinfachtes Deployment von Anwendungen können jedoch durchaus auch in Private Clouds realisiert werden.

## Rahmenbedingungen

Eine jüngere Literaturübersicht zeigt, dass der Datenschutz am häufigsten als die zentrale Herausforderung bei der Speicherung und Verarbeitung sensibler Gesundheitsdaten in der Cloud angesehen wird [15]. Eine aktuelle Umfrage belegt zudem eine gewisse Reserviertheit der Bevölkerung gegenüber Big-Data-Anwendungen und erhärtet die besondere Skepsis der Deutschen im internationalen Vergleich [18]. Zugriffe Dritter auf die Daten und der unbekannte Datenstandort in der Cloud wurden in dieser Reihenfolge auch am häufigsten als Sicherheitsbedenken von Firmen im Cloud-Scout-Report 2015 des Vereins „Deutschland sicher im Netz“ angegeben [19]. Auch der aktuelle Cloud-Monitor sieht Sicherheits- und Compliance-Bedenken in der Riskowahrnehmung nach wie vor an erster Stelle [2]. Insofern sind die datenschutzrechtlichen Vorgaben als kritische und ggf. auch prägende Rahmenbedingungen für das Cloud-Computing in der medizinischen Forschung anzusehen. Dies gilt umso mehr, als dass ein prominentes Anwendungsgebiet des Cloud-Computings in der Biomedizin die Speicherung und Verarbeitung genetischer Daten ist. Diese sind im Regelfall jedoch als personenbezogene Gesundheitsdaten anzusehen [20-22].

Das Vorliegen personenbezogener oder personenbeziehbarer Daten entscheidet auch über die Anwendbarkeit des Datenschutzrechts. Uneiniglich ist allerdings die recht-

liche Würdigung pseudonymer<sup>4</sup> oder verschlüsselter Daten hinsichtlich eines zu berücksichtigenden potentiellen Personenbezugs [vgl. 23, 24]. Eine Verschlüsselung der Daten in der Cloud ist insbesondere dann möglich und empfehlenswert, wenn die Cloud lediglich der Speicherung und Bereitstellung der Daten, nicht aber der Verarbeitung dient. Hierfür gibt es mittlerweile eine Reihe unterschiedlicher Dienste und Tools [25]. Die pseudonymisierte Speicherung der Daten in der Cloud erlaubt hingegen auch die Ausführung der allermeisten Verarbeitungsprozesse und schränkt doch die Risiken eines Missbrauchs schon deutlich ein [26]. Ein besonderer Ansatz verbirgt sich hinter dem Begriff der „Sealed Cloud“. Diese sieht nicht nur organisatorische sondern auch technische Maßnahmen zur Verhinderung einer Einsichtnahme von Daten durch den Dienstleister vor, auch wenn die Daten in der Cloud prozessiert und hierfür temporär entschlüsselt werden. Ein Metallkäfig um den betreffenden Server gibt den Zugriff auch für Wartungsarbeiten erst frei, wenn alle sensitiven Daten auf dem betreffenden System gelöscht wurden [27]. Eine einheitliche Bewertung dieser verschiedenen Schutzmaßnahmen durch die Datenschutzbeauftragten steht allerdings noch aus.

Datenschutzrechtliche Besonderheiten ergeben sich im Cloud-Szenario nur dann, wenn der Cloud-Nutzer und der Cloud-Anbieter nicht derselben juristischen Person zuzuordnen sind. Vollständig auszuschließen ist eine solche Konstellation nur im Falle einer Private Cloud, und das auch nur dann, wenn der Nutzer auch selbst Anbieter ist. In allen anderen Konstellationen ist zunächst die Verantwortlichkeit für die einzelnen Schritte der Erhebung, Verarbeitung und Speicherung der Daten zu klären. Das europäische (vgl. Art. 2d) Richtlinie 95/46/EG) und das deutsche Datenschutzrecht (vgl. § 3 Abs. 7 BDSG) knüpfen die rechtliche Verantwortlichkeit an die inhaltliche Verantwortlichkeit über die Entscheidung des Umgangs mit den Daten [24]. Demnach bleibt die rechtliche Verantwortlichkeit bei derjenigen Stelle, die primär über die Daten verfügt und diese rechtlich abgesichert zu Zwecken der Forschung verwenden darf. Die Voraussetzungen hierfür und unabhängig von einem Einsatz einer Cloud werden in den anderen Kapiteln dieses Reports dargelegt.

Das Datenschutzrecht sieht für eine solche Konstellation, bei der die Verantwortung vollständig bei einer Stelle verbleibt, die Möglichkeit der Datenverarbeitung im Auftrag nach § 11 BDSG vor. In diesem Fall wird der Cloud-Nutzer zum Auftraggeber und der Cloud-Anbieter zum Auftragnehmer. Ein solcher Auftrag ist nach § 11 Abs. 2 BDSG schriftlich zu erteilen und muss u. a. Festlegungen zu Umfang, Dauer, Art und Zweck der vorgesehenen Verarbeitung oder Nutzung von Daten, zur Art der Daten und dem Kreis der Betroffenen, über die technischen und organisatorischen Maßnahmen

4 Siehe hierzu auch den Abschnitt zu den rechtlichen Rahmenbedingungen im Kapitel zur Sekundärnutzung klinischer Daten im vorliegenden Report.

zum Schutz der Daten, zu Kontroll- und Meldepflichten und zu möglichen Unterauftragsverhältnissen enthalten.

Das Grundprinzip dieser Regelungen kann man sich verdeutlichen, wenn man davon ausgeht, dass ein betroffener Patient, dessen Daten verarbeitet werden, unabhängig von einer Datenverarbeitung im Auftrag gegenüber dem Verantwortlichen für die Datenverarbeitung, in diesem Falle also dem Auftraggeber, immer die gleichen Rechte einfordern können muss. Dies gilt für die zugesicherte Sicherheit der Daten alle auch sonst geltenden Patientenrechte.

Eine im Kontext des Cloud-Computings relevante Frage betrifft die notwendige Transparenz der Datenverarbeitung für den Auftraggeber im Sinne einer umfassenden Informiertheit. Die Datenschützer setzen eine solche Transparenz voraus [24, 28], so dass z. B. Unterauftragsverhältnisse nicht nur gestattet werden müssen, sondern der Auftragnehmer auch noch zu verpflichten ist, sämtliche Unterauftragnehmer dem Auftraggeber gegenüber abschließend zu benennen [24, S. 9].

Tatsächlich sehen die Autoren der „Orientierungshilfe – Cloud Computing“ der Arbeitskreise Technik und Medien der Konferenz der Datenschutzbeauftragten des Bundes und der Länder weitreichende Kontrollpflichten für den Cloud-Nutzer (dort Cloud-Anwender genannt) vor [24]. Zwar wird eingeräumt, dass eine Vor-Ort-Prüfung nicht immer möglich sein wird, aber eine bloße Zusicherung des Cloud-Anbieters, dass alle datenschutzrechtlichen Voraussetzungen erfüllt werden, erscheint nicht ausreichend. Daher wird vorgeschlagen, dass der Cloud-Anbieter sich einem Zertifizierungs- bzw. Gütesiegelverfahren zu Fragen des Datenschutzes und der Datensicherheit bei einer unabhängigen Prüfstelle unterwirft. Allerdings reicht das bloße Vorliegen eines Gütesiegels oder Zertifikats nicht aus, der Cloud-Nutzer muss sich auch davon überzeugen, dass alle inhaltlich relevanten Aspekte seiner Kontrollpflichten von den jeweiligen Prüfverfahren abgedeckt sind [24, S. 10f]. Diese Ansprüche der Datenschützer an Kontrollmaßnahmen des Cloud-Nutzers gelten gleichermaßen für die Überprüfung von Unterauftragnehmern [24, S. 11f]. Ein allgemein anerkanntes Zertifikat steht noch nicht zur Verfügung [vgl. 29]. Wichtige Vorarbeiten wurden vom BMWi im Rahmen des Trusted-Cloud-Programms angestoßen [30, 31].

Grundsätzlich erscheint aus der Perspektive des Datenschutzes eine möglichst transparente Cloud wünschenswert, in der der Nutzer genau den Ort und die Art der Verarbeitung seiner Daten bestimmen kann [32, S. 683]. Allerdings ist diese Art der Transparenz, die man mit „Durchsichtigkeit“ übersetzen könnte, das Gegenteil jener Transparenz, von der im IT-Umfeld häufig im Sinne einer „Unsichtbarkeit“ gesprochen wird. Wenn von

einer transparenten Nutzung von Cloud-Ressourcen die Rede ist, wird üblicherweise davon ausgegangen, dass der Nutzer sich gerade um die Details des Wo und Wie der Datenverarbeitung in der Cloud nicht mehr kümmern muss und kann. Diese Form der Transparenz im Sinne einer „Unsichtbarkeit“ und die damit verbundene Möglichkeit des Auftraggebers, sich gerade um solche Detailfragen nicht mehr kümmern zu müssen, wird jedoch von Datenschützern als unzulässig eingestuft [32, S. 683].

Die hier dargestellte Komplexität der Rahmenbedingungen ist entsprechend auch bei der Vertragsgestaltung zur Umsetzung einer Auslagerung von IT-Services zu berücksichtigen. Dabei ist eine Vielzahl technischer, organisatorischer und rechtlicher Aspekte zu beachten [vgl. 33, 34].

### **Nutzung internationaler Cloud-Angebote**

Clouds werden heute im kommerziellen Umfeld international und damit grenzüberschreitend angeboten [1, 2]. Gerade Kosteneffekte führen zu einer Zentralisierung der für ein Cloud-Angebot nötigen Hardware und einer Adressierung eines möglichst großen und damit notwendigerweise internationalen Markts. Im Folgenden wird beleuchtet, welche Folgen sich für einen deutschen Nutzer eines Cloud-Services ergeben, wenn der Cloud-Anbieter seinen Sitz nicht in Deutschland hat, bzw. wenn die Daten in der Cloud den deutschen Rechtsraum verlassen. Für eine internationale Perspektive wäre auch die umgekehrte Sichtweise von Interesse, also wenn deutsche Cloud-Angebote aus dem Ausland heraus genutzt werden. Die rechtlichen Bedingungen für diese „umgedrehte“ Perspektive werden hier jedoch nicht weiter untersucht.

Aufgrund der Harmonisierung des Datenschutzrechts in der EU entlang der Richtlinie 95/46/EG gelten für die Nutzung eines physikalisch und rechtlich innerhalb eines Mitgliedslands der EU oder des Europäischen Wirtschaftsraums (EWR) aufgestellten Cloud-Angebots keine zusätzlichen Regeln oder Voraussetzungen [24, S. 14], auch wenn einzelne Datenschützer auf die geringeren effektiven Kontrollmöglichkeiten oder einen ggf. fehlenden Beschlagnahmeschutz von Cloud-Angeboten außerhalb des deutschen Rechtsraums hinweisen [32]. Diese rechtliche Grenzziehung ist im deutschen Datenschutzrecht auf nationaler Ebene in § 3 (8) Satz 2 BDSG in Form der Definition „Dritter“ normiert. „Dritte“ sind demnach alle Personen oder Stellen außerhalb der verantwortlichen Stelle und ausgenommen die betroffene Person selbst. Interessanterweise gelten nun aber Personen oder Stellen, die innerhalb der EU (bzw. des EWR) Daten im Auftrag verarbeiten, nicht als „Dritte“.

Für die Übermittlung von personenbezogenen Daten an „Dritte“ bzw. Empfänger in Drittstaaten gilt somit nicht die Privilegierung der Regelungen zur Auftragsdatenverarbeitung. In diesen Fällen übernehmen die Datenempfänger in den Drittstaaten auch einen Teil der Verantwortung für die Datenverarbeitung. Diese Übertragung eines Teils der Verantwortung, Weichert spricht von einer Funktionsübertragung [32, S. 682], benötigt jedoch eine eigenständige Rechtsgrundlage. Zudem erscheint die Übernahme einer Verantwortung aus der Sicht der Anbieter einer Cloud problematisch, da ja gerade eine inhaltliche Kontrolle der Daten und ihrer Verarbeitung durch den Anbieter im Regelfall weder gewünscht noch möglich ist.

Häufig wird als Rechtsgrundlage für eine Datenübermittlung an einen Empfänger in einem Drittstaat § 28 (1) Satz 1 Nr. 2 BDSG genannt, in dem ein Erfordernis der Datenübermittlung für die Wahrung berechtigter Interessen der verantwortlichen Stelle, wenn gleichzeitig schutzwürdige Interessen der Betroffenen nicht überwiegen, vorausgesetzt wird. Allerdings ist diese Rechtsgrundlage für besonders schutzwürdige Daten gemäß § 3 (9) BDSG als Möglichkeit ausgeschlossen [24, S. 16f]. Eine relevante Rechtsgrundlage für eine Datenübermittlung klinischer Daten findet sich hingegen im Arzneimittelrecht: Nach § 40 (2a) AMG sind die personenbezogenen Daten der Betroffenen pseudonymisiert an den Sponsor zum Zwecke der wissenschaftlichen Auswertung weiterzugeben.

Da die USA Sitz vieler bedeutender Anbieter von IT-Services und zudem wichtiger Forschungsstandort ist, stellt sich die Frage nach der Nutzbarkeit dort lokalisierter Cloud-Angebote. Allgemeiner formuliert, ist dies die Frage nach den rechtlichen Rahmenbedingungen für Clouds in sogenannten Drittstaaten, die nicht den Regeln der EU oder des EWR unterliegen. Gemäß Artikel 25 (6) der Richtlinie 95/46/EG kann die Europäische Kommission in einem ausführlich geregelten Verfahren festlegen, in welchen Drittstaaten ein adäquater und vergleichbarer gesetzlicher Datenschutz umgesetzt ist. Länder außerhalb der EU und des EWR, für die es keine solche Feststellung der Europäischen Kommission gibt, wie auch die USA, sind demnach gesondert zu betrachten.

Bis zum Oktober 2015 gab es für US-amerikanische Firmen die Möglichkeit, sich im Rahmen einer Selbstverpflichtung gegenüber dem US-Handelsministerium nach den Safe-Harbor-Prinzipien für ein Datenschutzniveau nach EU-Vorgaben zu qualifizieren. Der Europäische Gerichtshof hat im Oktober 2015 jedoch die Anwendung dieser Regelung, die auf eine Entscheidung der Europäischen Kommission aus dem Jahre 2000 zurückging, für ungültig erklärt, u. a. da vor dem Hintergrund der Enthüllungen Edward Snowdens in den USA von keinem ausreichenden Schutz der Daten vor staatlichen Stellen wie der NSA ausgegangen werden kann [35].

Als alternative rechtliche Grundlage einer Übermittlung personenbezogener Daten in die USA kam bisher die Nutzung hierfür definierter EU-Standardvertragsklauseln oder eine entsprechende Selbstverpflichtung in sogenannten Corporate Binding Rules zur Anwendung. Beide Verfahren sind von der jüngsten EU-Rechtsprechung nicht direkt berührt. Die Begründung des EuGH zum Safe-Harbor-Urteil macht jedoch deutlich, dass auch diese Verfahren grundsätzlich in Zweifel gezogen werden können. Zudem hat der EuGH in seinem Urteil den Prüfauftrag der zuständigen Aufsichtsbehörden deutlich heraus gestellt, der auch unabhängig von der Anwendung der Safe-Harbor-Vereinbarung oder analoger Regelungen gilt [35]. Die zuständige Vereinigung der Datenschutzbeauftragten auf europäischer Ebene hat bereits dargelegt, dass sie diesen Prüfauftrag ernst nimmt und auf dieser Basis auch Datenübermittlungen auf Basis der Alternativregelungen zum Safe-Harbor-Abkommen kritisch unter die Lupe nehmen wird [36].

Als Nachfolgeregelung zu Safe Harbor wurde 2016 zwischen der EU und den Vereinigten Staaten von Amerika der EU-US Privacy Shield ausgehandelt und am 12.6.2016 von der EU-Kommission als angemessen in Bezug auf die Einhaltung eines zur EU adäquaten Datenschutzniveaus beurteilt. Auch diese Regelung enthält eine Selbstverpflichtungskomponente, die von US-amerikanischen Firmen genutzt werden kann. Diese wird jedoch, anders als noch bei Safe Harbor, ergänzt um staatliche Garantien und Aufsichtsfunktionen. Mit dieser Regelung und dem zugehörigen Angemessenheitsbeschluss der EU-Kommission besteht eine neue rechtliche Grundlage für die Übermittlung personenbezogener Daten an entsprechende Firmen in den USA, die dem EU-US Privacy Shield unterliegen. Von vielen Seiten wird diese Regelung allerdings kritisch gesehen. Insbesondere werden die staatlichen Garantien seitens der USA angesichts der nicht eingeschränkten staatlichen Zugriffsbefugnisse im Rahmen der Terror-Bekämpfung für nicht ausreichend gehalten [so z. B. 37, 38]. Insofern ist davon auszugehen, dass auch gegen diese Regelung geklagt wird und ein aufhebendes Urteil des EuGH wie zu Safe Harbor dann nicht ausgeschlossen werden kann.

Als zulässige Rechtsgrundlage für eine Übermittlung personenbezogener Daten an Stellen, bei denen ein ausreichendes Datenschutzniveau nicht vorausgesetzt werden kann, bleibt grundsätzlich auch die in § 4c Abs. 1 Nr. 1 BDSG vorgesehene Einwilligung bestehen. Dies Verfahren setzt allerdings voraus, dass der Betroffene explizit auf die Übermittlung und das nicht ausreichende Datenschutzniveau beim Empfänger hingewiesen wird. Es ist allerdings bereits angemerkt worden, dass die US-amerikanischen Ermächtigungsklauseln für die Geheimdienste [vgl. 39] nicht vereinbar mit dem für die Einwilligung zentralen Prinzip der Zweckbindung sein könnten [40].

## **Grenzen einer Datenverarbeitung im Auftrag**

Wie bereits im Kapitel zur Sekundärnutzung klinischer Daten im vorliegenden Report dargestellt, gilt mit der ärztlichen Schweigepflicht für die Verarbeitung der im Rahmen der Behandlung erhobenen Daten ein zur datenschutzrechtlichen Zweckbindung zusätzliches und von dieser unabhängiges zweites Schutzprinzip. Dieses betrifft zwar nicht die interne Forschung durch die Behandler selbst, wirkt aber immer dann als zusätzliche Schranke, wenn die Daten die behandelnde Einrichtung verlassen sollen, wie dies beim Einsatz von Clouds regelmäßig der Fall sein wird. Die datenschutzrechtliche Privilegierung im Rahmen einer Datenverarbeitung im Auftrag nach § 11 BDSG kann aber gemäß § 1 Abs. 3 Satz 2 BDSG gesetzliche Geheimhaltungspflichten und Berufsgeheimnisse wie die ärztliche Schweigepflicht, in den Berufsordnungen und in § 203 StGB normiert, nicht aufheben [41, 42].

Somit können Behandlungsdaten nur auf Basis einer Einwilligung oder einer spezialgesetzlich geregelten Erlaubnis zur Übermittlung oder Auftragsdatenverarbeitung in die Cloud gegeben werden. Der von der TMF beauftragte Rechtsgutachter Schneider kommt nach einer umfassenden Sichtung aller in Frage kommenden Spezialgesetze zu dem Schluss, dass in Deutschland nur in den Stadtstaaten Bremen und Hamburg in den jeweiligen Landeskrankenhausgesetzen eine Auftragsdatenverarbeitung normiert wird, die weitgehend den Maßstäben der Regelung nach § 11 BDSG entspricht. In den Bundesländern Schleswig-Holstein, Niedersachsen, Sachsen-Anhalt und Brandenburg gibt es hingegen keinerlei spezialgesetzliche Erlaubnis zur Auftragsdatenverarbeitung. Alle übrigen Bundesländer normieren in ihren Landeskrankenhausgesetzen (bzw. dem Gesundheitsdatenschutzgesetz in NRW) eine im Vergleich zum BDSG deutlich eingeschränkte und an zusätzliche Auflagen gebundene Auftragsdatenverarbeitung. Für alle Einrichtungen, für die lediglich das BDSG anwendbar ist, wie beispielsweise Arztpraxen, fehlt hingegen eine Rechtsgrundlage zur Auftragsdatenverarbeitung, die gleichzeitig als Offenbarungsbefugnis nach § 203 StGB angesehen werden könnte [42].

## **Ausblick auf die EU-Datenschutzgrundverordnung**

Im Frühjahr 2016 wurde nach vierjähriger Verhandlung die EU-Datenschutzgrundverordnung (EU-DSG-VO) verabschiedet, die ab Mai 2018 anwendbar sein wird. Damit wird das grundsätzliche Niveau des Datenschutzes in der EU neu definiert [43]. Entsprechend sind auch Vereinbarungen wie die Privacy-Shield-Regelung dann nach Art. 45 EU-DSG-VO zu überprüfen und ggf. neu auszurichten.<sup>5</sup> Die Nutzung von EU-Stan-

---

<sup>5</sup> Bis zu einer Überprüfung bleibt die bisherige Angemessenheitsentscheidung nach Art. 45 Abs. 9 gültig.

dardvertragsklauseln und die Einholung von Einwilligungen bleiben zudem auch nach der EU-DSG-VO weiterhin zulässige Alternativen.

Zudem definiert die EU-DSG-VO unmittelbarer als die bisherige Richtlinie 95/46/EG auch das jeweils national anwendbare Datenschutzrecht. Ein solcher Harmonisierungseffekt wäre gerade für die heterogene Rechtslage zum Datenschutz für die biomedizinische Forschung in Deutschland sehr wünschenswert. Allerdings lässt die EU-DSG-VO den Nationalstaaten in Art. 9 Abs. 4 gerade in Bezug auf Gesundheitsdaten viel Spielraum für eigene und ggf. auch engere Regelungen. Derzeit ist noch sehr offen, wie die Gesetzgebung im Bund und in den Ländern mit diesem Spielraum umgehen wird. Es ist leider nicht ausgeschlossen, dass auch mit der EU-DSG-VO viel mehr „beim Alten bleiben“ wird, als viele vorhergesagt und eigentlich auch gehofft hatten.

## Aktueller Bestand

Wie bereits in dem Abschnitt zum Anwendungsbereich dargestellt wurde, gibt es mit den Kriterien des NIST eine enge Definition des Einsatzes einer Cloud [5]. Allerdings wurde auch dort schon darauf hingewiesen, dass nach diesen Maßstäben viele Projekte, die unter dem Stichwort „Cloud“ publiziert werden, im engeren Sinne nicht als Cloud-Computing angesehen werden können [15]. Vor diesem Hintergrund kann auch nicht ausgeschlossen werden, dass die Überschrift „Cloud-Computing“ in einigen Fällen aus Marketinggründen verwendet wird, oder der Adressierung eines passenden Förderprogramms dient. Diese methodischen Probleme zusammen mit dem Fehlen entsprechend detaillierter und sorgfältiger Bestandsübersichten erschweren eine realistische Einschätzung der aktuellen Relevanz des Cloud-Computings in der medizinischen Forschung.

Auch wenn es an konkreten und nach hart definierten und dokumentierten Kriterien erhobenen Zahlen fehlt, lassen sich doch einige Trends ausmachen. So kann im Bereich der Speicherung und Verarbeitung molekularbiologischer Daten entsprechend einiger Veröffentlichungen von einem steigenden Einsatz von Clouds ausgegangen werden [4, 6-8]. Wichtige Gründe hierfür sind die enorme und rasch zunehmende Menge der zu verarbeitenden Daten sowie die aufwändigen und sich dynamisch entwickelnden Verarbeitungsprozesse.

Dieselben Übersichtsarbeiten belegen auch einen weiteren Trend, der darin besteht, dass die Auslagerung von IT-Services nach dem Cloud-Paradigma im internationalen Umfeld und hier insbesondere außerhalb der EU häufiger als in der EU oder gar in

Deutschland anzutreffen ist. Es ist mehr als naheliegend, hierfür die engeren rechtlichen Rahmenbedingungen in der EU und insbesondere in Deutschland verantwortlich zu machen. Inwiefern innovative und aus Deutschland stammende technische Entwicklungen [27, 44] zu einer größeren Akzeptanz des Cloud-Einsatzes für Gesundheitsdaten führen, muss allerdings derzeit noch verhalten skeptisch beurteilt werden.

Ein dritter Trend hat mit dem zunehmenden Einsatz mobiler Devices für die Datenerfassung in klinischen Forschungsprojekten zu tun. Die Verarbeitung der von Smartphones und den daran gekoppelten Sensoren erfassten multidimensionalen Daten erfolgt schon heute fast ausschließlich auf zentralen und häufig dynamisch verwalteten Servern. Auch diese Entwicklung spricht daher für einen Anstieg medizinischer Forschungsdaten in der Cloud.

Ein hemmender Faktor für die Verbreitung der Cloud-Nutzung in der öffentlich geförderten Forschung ist sicher noch in den einschlägigen Förderrichtlinien zu sehen, die Investitions- gegenüber Betriebskosten bevorzugen. Erstere sind häufig förderfähig, letztere hingegen im Regelfall nicht. Eine aktuelle Stellungnahme der Kommission für IT-Infrastruktur für 2016–2020 der DFG zeigt allerdings, dass hier ein Umdenken bei den Förderorganisationen begonnen hat. Künftig sollten Förderer die Gesamtkosten (Total Cost of Ownership) von IT-Diensten stärker berücksichtigen, um so die Nutzung von Cloud-Diensten mit Hardware-Anschaffungen und entsprechendem Eigenbetrieb vergleichbar und ggf. förderbar zu machen [45].

Auch die Europäische Kommission sieht Handlungsbedarf und hat eine Experten-Gruppe ein erstes Konzeptpapier für eine European Open Science Cloud (EOSC) erarbeiten lassen [46]. Darin wird insbesondere auf die Notwendigkeit der sorgfältigen Aufbereitung, Bewahrung und Verfügbarmachung der enormen Datenmengen in der heutigen Wissenschaft hingewiesen. Auch wenn der Begriff der Cloud hier eher als Metapher für die vollständige Verfügbarkeit aller wissenschaftlich relevanten Daten unabhängig vom Standort Verwendung findet und das Konzeptpapier entsprechend nicht näher auf die technischen Details des Cloud-Computings eingeht, wird doch ein umfangreiches Förderprogramm für entsprechende Cloud-Infrastrukturen in der EU vorgeschlagen [46].

Auf deutscher Seite ist diese Initiative insbesondere von der Helmholtz Gemeinschaft aufgegriffen worden, die sich hier mit ihren Großforschungseinrichtungen als Ansprechpartner sieht und aktuell ein Positionspapier zum offenen Umgang mit Forschungsdaten verabschiedet hat [47]. Für den Bereich der Lebenswissenschaften ist in diesem Kontext auch auf die stark vom Heidelberger DKFZ geprägte Initiative zur Errichtung einer Applied & Translational Genomics Cloud (ATGC) hinzuweisen [48],

der sich mittlerweile auch das Deutsche Netzwerk für Bioinformatik-Infrastruktur (de. NBI) angeschlossen hat.<sup>6</sup>

Vor dem Hintergrund dieser verschiedenen Entwicklungen und Initiativen ist künftig von einem starken Einfluss von Cloud-Infrastrukturen auch auf den Bereich der akademisch geprägten und öffentlich geförderten Forschung auszugehen. Damit ist selbst dann zu rechnen, wenn nicht alle Ankündigungen wahr werden und einige der genannten Initiativen womöglich am Ende doch im Sande verlaufen.

## Bewertung und Handlungsbedarf

Grundsätzlich erscheint eine Weiterentwicklung der IT-Infrastrukturen in der medizinischen Forschung hin zu einer weitergehenden Service-Orientierung wünschenswert.<sup>7</sup> Hierfür spricht die in Deutschland auf sehr viele Standorte verteilte Forschungsaktivität. Dabei werden an einem einzelnen Standort nicht selten auch noch von mehreren Abteilungen unabhängig voneinander Infrastrukturen für die Forschung aufgebaut. Es ist aber nicht davon auszugehen, dass alle diese Einrichtungen und Abteilungen, zumal bei stark begrenzten Budgets, eine allen Qualitätsanforderungen genügende Forschungs-IT entwickeln und betreiben können. Eine Reihe deutscher Forscher kommt in einem White Paper zur Entwicklung einer deutschen Cloud für „Applied & Translational Genomics“ sogar zu dem Schluss, dass keine einzige deutsche Universität oder Forschungseinrichtung derzeit über die notwendige Infrastruktur verfügt, um mit den absehbaren Datenmengen im Omics-Bereich Analysen durchzuführen und eine gesicherte Speicherung sowie langfristigen Zugriff auf die Daten zu gewährleisten [48]. Die in Deutschland stark vertretene Verbundforschung, die nicht zuletzt auf ein Gesundheitssystem zurückgeht, das durch eine räumlich sehr gleichmäßige Verteilung von Gesundheitseinrichtungen charakterisiert ist, ist im Ergebnis auf eine zunehmende Zentralisierung von Infrastrukturen angewiesen.

Im Zuge des allgemeinen Trends zum Cloud-Computing findet derzeit eine dynamische technische Entwicklung von Schnittstellen, Verwaltungsstandards- und Tools für die weitgehend automatisierbare Bereitstellung von Services auf Infrastruktur-, Plattform-, Anwendungs- und Software-Ebene statt. Auch wenn die Nutzung kommerzieller Cloud-Angebote in der medizinischen Forschung nicht immer mit der versprochenen Kosteneinsparung einher gehen oder sich rechtlich einfach umsetzen lassen wird, sollte

---

6 Hierzu wurde im April 2016 ein gemeinsamer Workshop durchgeführt, siehe [www.applied-translational-genomics-cloud.de/joomla/index.php/en/workshop](http://www.applied-translational-genomics-cloud.de/joomla/index.php/en/workshop).

7 Zu diesem Schluss kam auch ein 2010 von der TMF durchgeführter Workshop zum IT-Service-Management, siehe [www.tmf-ev.de/news/738](http://www.tmf-ev.de/news/738).

diese technische Entwicklung doch sorgfältig beobachtet werden. Es ist davon auszugehen, dass viele der technischen Neuerungen aus dem Cloud-Bereich künftig auch für eine allgemeine Service-Orientierung der IT einsetzbar sein oder diese sogar weitgehend prägen werden. Von dieser Entwicklung können somit auch Forschungseinrichtungen profitieren, die zentrale IT-Services im Forschungsverbund oder auch für mehrere Forschungsverbünde anbieten. Beispielhaft hierfür sei die Nutzung der Bereitstellung von Software und Services auf Basis von Containern nach dem Docker-Standard genannt [49, 50]. Ein TMF-Workshop hierzu stieß auf große Resonanz<sup>8</sup> und aktuell wird die Nutzbarkeit dieser Technologie im akademischen Umfeld im Rahmen eines TMF-finanzierten Projekts evaluiert [51].

Im Idealfall können auf einer solchen Grundlage auch „kleine“ und mit kurzer Laufzeit angesetzte Forschungsprojekte kosteneffizient umgesetzt werden, in dem sie ohne großen administrativen Aufwand auszulösen, qualitativ hochwertige Ressourcen und Services buchen und nutzen können. Neben den technischen Standards wird allerdings auch ein effizientes Vertragsmanagement ein entscheidender Akzeptanzfaktor sein. Dies schließt ggf. auch die Bereitstellung geeigneter und rechtlich abgesicherter Einwilligungensformulare mit ein. Um Forschungseinrichtungen und Projekten die passende Auswahl der für sie geeigneten IT-Services oder Cloud-Angebote zu ermöglichen, sollte ein Kriterienkatalog oder eine Checkliste entwickelt werden. Diese sollte einen Abgleich der typischen Anforderungen mit den angebotenen Leistungen unterstützen und dabei auch eine realistische Kostenschätzung für den im Regelfall erforderlichen mehrjährigen Betrieb ermöglichen. Für eine Risiko-Bewertung und die sich daraus ergebenden Anforderungen an die IT-Sicherheit liegen bereits Vorarbeiten vor [52, 53].

Grundsätzlich bietet der europäische Rechtsraum einen geeigneten Rahmen für die Nutzung von Clouds. Die im Frühjahr 2016 verabschiedete Fassung einer neuen europäischen Datenschutzgrundverordnung wird aller Voraussicht nach für eine weitere Vereinheitlichung der datenschutzrechtlichen Bestimmungen sorgen, auch wenn gerade im Umgang mit Gesundheitsdaten womöglich auch weiterhin nationale Sonderbestimmungen Gültigkeit behalten. Es gibt zudem Hinweise, dass das hiesige Datenschutzniveau als Standortfaktor für Clouds zunehmend Berücksichtigung findet [2]. Vor diesem Hintergrund ist auch die aus Heidelberg stammende und mittlerweile vom de.NBI aufgegriffene Initiative zur Entwicklung einer deutschen Cloud für Applied & Translational Genomics zu begrüßen [48]. Auch wenn noch nicht klar erkennbar ist, wann und in welcher Form künftig eine solche deutsche Genomics-Cloud angeboten

---

8 Siehe [www.tmf-ev.de/news/2809](http://www.tmf-ev.de/news/2809).

wird, ist es doch verdientvoll, wenn sich bereits jetzt akademische Einrichtungen und kommerzielle Partner gemeinsam zu den Anforderungen an eine solche Cloud sowie den relevanten Rahmenbedingungen abstimmen.

## Literatur

1. Leimbach, T., Bachlechner, D. *TA-Vorstudie: Big Data in der Cloud*. 2014. Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag: TAB-Hintergrundpapier, <http://www.tab-beim-bundestag.de/de/aktuelles/20150421.html> (Abruf: 2015-04-24).
2. KPMG *Cloud-Monitor 2016. Cloud-Computing in Deutschland – Status quo und Perspektiven*. 2016. Bitkom Research GmbH im Auftrag der KPMG AG Wirtschaftsprüfungsgesellschaft, <http://hub.klardenker.kpmg.de/cloud-monitor-2016> (Abruf: 2016-10-24).
3. DFG *Cloud-Dienste - Addendum zu den Empfehlungen der Kommission für IT Infrastruktur 2011-2015*. 2014. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/foerderung/programme/wgi/addendum\\_cloud\\_dienste\\_kfr\\_2014.pdf](http://www.dfg.de/download/pdf/foerderung/programme/wgi/addendum_cloud_dienste_kfr_2014.pdf) (Abruf: 2016-12-01).
4. Eugster, M.J.A., Schmid, M., Binder, H., Schmidberger, M., *Grid and Cloud Computing Methods in Biomedical Research*. *Methods of Information in Medicine*, 2013. **52**: S. 62-64.
5. Mell, P., Grance, T. *The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology*. 2011. National Institute of Standards and Technology, NIST Special Publication 800-145.
6. Marx, V., *Genomics in the clouds*. *Nat Methods*, 2013. **10**(10): S. 941-945.
7. Marx, V., *Biology: The big challenges of big data*. *Nature*, 2013. **498**(7453): S. 255-260.
8. Stein, L.D., *The case for cloud computing in genome informatics*. *Genome Biol*, 2010. **11**(5): S. 207.
9. Grütz, R., Franke, T., Dickmann, F., *Concept for Preservation and Reuse of Genome and Biomedical Imaging Research Data*. *Studies in Health Technology and Informatics - Medinfo 2013*, 2013. **192**: S. 999.
10. Bernau, C., Boulesteix, A.-L., Knaus, J., *Application of Microarray Analysis on Computer Cluster and Cloud Platforms*. *Methods of Information in Medicine*, 2013. **52**: S. 65-71.
11. Pütz, B., Kam-Thong, T., Karbalai, N., Altmann, A., Müller-Myshok, B., *Cost-effective GPU-Grid for Genome-wide epistasis calculations*. *Methods of Information in Medicine*, 2013. **52**: S. 91-95.

12. Piotrowski, M., McGilvary, G.A., Sloan, T.M. et al., *Exploiting Parallel R in the cloud with SPRINT*. Methods of Information in medicine, 2013. **52**: S. 80-90.
13. Christoph, J., Griebel, L., Leb, I. et al., *Secure Secondary Use of Clinical Data with Cloud-based NLP Services Towards a Highly Scalable Research Infrastructure*. Methods of Informatics in Medicine, 2015. **54**(3 2015): S. 276-282.
14. Ohmann, C., Canham, S., Danielyan, E., Robertshaw, S., Legré, Y., Clivio, L., Demotes, J., *Cloud computing and clinical trials: report from an ECRIN workshop*. Trials, 2015. **16**: S. 318-321.
15. Griebel, L., Prokosch, H.U., Köpcke, F. et al., *A scoping review of cloud computing in healthcare*. BMC Med Inform Decis Mak, 2015. **15**: S. 17.
16. Chatziastros, A., Drepper, J., Semler, S.C., *BIG DATA und Healthcare Analytics – datenschutzrechtliche Herausforderungen*. mdi, 2014. **2014**(2): S. 53-57.
17. Knaus, J., Hieke, S., Binder, H., Schwarzer, G., *Costs of Cloud Computing for a Biometry Department - A Case Study*. Methods of Information in Medicine, 2013. **52**: S. 72-79.
18. Vodafone *Big Data. A European Survey on the Opportunities and Risks of Data Analytics*. 2016. Vodafone Institute for Society and Communications, <http://www.vodafone-institut.de/wp-content/uploads/2016/01/VodafoneInstitute-Survey-BigData-en.pdf>.
19. DsiN *DsiN Cloud-Scout Report 2015. Cloud Computing – Kleine und mittlere Unternehmen in Europa*. 2015. Deutschland sicher im Netz e.V., <https://www.sicher-im-netz.de/downloads/dsin-cloud-scout-report-2015> (Abruf: 2016-01-25).
20. Rodriguez, L.L., Brooks, L.D., Greenberg, J.H., Green, E.D., *The Complexities of Genomic Identifiability*. Science, 2013. **339**(6117): S. 275-276.
21. 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.
22. Dove, E.S., Joly, Y., Tasse, A.M. et al., *Genomic cloud computing: legal and ethical points to consider*. Eur J Hum Genet, 2015. **23**(10): S. 1271-1278.
23. Hon, W.K., Millard, C., Walden, I., *The problem of ‚personal data‘ in cloud computing: what information is regulated? - the cloud of unknowing*. International Data Privacy Law, 2011. **1**(4): S. 211-228.
24. DSK *Orientierungshilfe – Cloud Computing*. 2014. Arbeitskreise Technik und Medien der Konferenz der Datenschutzbeauftragten des Bundes und der Länder sowie der Arbeitsgruppe Internationaler Datenverkehr des Düsseldorfer Kreises, [http://www.datenschutz-bayern.de/technik/orient/oh\\_cloud.pdf](http://www.datenschutz-bayern.de/technik/orient/oh_cloud.pdf) (Abruf: 2015-01-13).
25. Schmech, K., *Wolkenschlösser: Verschlüsselungs-Gateways für die Cloud*. ix, 2015. **2015**(12): S. 106-110.

26. Schüler, P., *Geheimnis-Verwalter. Das Eperi Gateway schützt Cloud-Daten.* c't, 2016. **2016**(4): S. 174-177.
27. Adis, W., *Gesiegelt. Kurz erklärt: Sealed Cloud.* ix, 2015. **2015**(8): S. 117.
28. Artikel-29-WP *Opinion 05/2012 on Cloud Computing.* 2012. Article 29 Data Protection Working Party, [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp196\\_en.pdf](http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf) (Abruf: 2014-02-05).
29. Schulzki-Haddouti, C., *Zwischen Orientierungshilfe und Marketing-Gag. Zertifizierungen für Cloud-Anbieter.* c't, 2016. **2016**(11): S. 146-151.
30. Kompetenzzentrum-Trusted-Cloud *Trusted Cloud-Datenschutzprofil für Cloud-Dienste (TCDP) – Version 0.9.* 2015. Kompetenzzentrum Trusted Cloud, [http://trustedcloud.pt-dlr.de/media/content/Publikation\\_TCDP.pdf](http://trustedcloud.pt-dlr.de/media/content/Publikation_TCDP.pdf) (Abruf: 2016-01-24).
31. Kompetenzzentrum-Trusted-Cloud *Thesenpapier – Eckpunkte eines Zertifizierungsverfahrens für Cloud-Dienste.* 2015. Kompetenzzentrum Trusted Cloud, [http://trustedcloud.pt-dlr.de/media/content/150407\\_Nr\\_12\\_Zertifizierung\\_gesamt\\_RZ\\_Ansicht\\_EZ.pdf](http://trustedcloud.pt-dlr.de/media/content/150407_Nr_12_Zertifizierung_gesamt_RZ_Ansicht_EZ.pdf) (Abruf: 2016-01-24).
32. Weichert, T., *Cloud Computing und Datenschutz.* Datenschutz und Datensicherheit, 2010. **2010**(10): S. 679-687.
33. Haar, T., *Mietsache. Vertragliche Gestaltung bei SaaS und Cloud-Computing.* ix, 2014. **2014**(10): S. 100-101.
34. Kompetenzzentrum-Trusted-Cloud *Leitfaden – Vertragsgestaltung beim Cloud Computing.* 2015. Kompetenzzentrum Trusted Cloud, [http://trustedcloud.pt-dlr.de/media/content/140317\\_Vertragsleitfaden\\_gesamt\\_RZ\\_Ansicht.pdf](http://trustedcloud.pt-dlr.de/media/content/140317_Vertragsleitfaden_gesamt_RZ_Ansicht.pdf) (Abruf: 2016-01-24).
35. EuGH *Urteil des Gerichtshofes (Große Kammer) vom 6. Oktober 2015. Maximilian Schrems gegen Data Protection Commissioner. Aktenzeichen C-362/14.* 2015. Europäischer Gerichtshof.
36. Artikel-29-WP *Statement on the implementation of the judgement of the Court of Justice of the European Union of 6 October 2015 in the Maximilian Schrems vs. Data Protection Commissioner case (C-362-14).* 2015. Article 29 Data Protection Working Party, [http://ec.europa.eu/justice/data-protection/article-29/press-material/press-release/art29\\_press\\_material/2015/20151016\\_wp29\\_statement\\_on\\_schrems\\_judgement.pdf](http://ec.europa.eu/justice/data-protection/article-29/press-material/press-release/art29_press_material/2015/20151016_wp29_statement_on_schrems_judgement.pdf).
37. Artikel-29-WP *Opinion 01/2016 on the EU – U.S. Privacy Shield draft adequacy decision.* 2016. Article 29 Data Protection Working Party, [http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2016/wp238\\_en.pdf](http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2016/wp238_en.pdf) (Abruf: 2016-10-23).
38. Schaar, P., *Balance-Akt. Wie Privacy Shield EU-Bürger schützen soll.* c't, 2016. **2016**(7): S. 66-69.

39. Boehm, F. *A comparison between US and EU data protection legislation for law enforcement purposes. Study for the LIBE Committee*. 2015. Directorate General for Internal Policies. Policy Department C: Citizens' Rights and Constitutional Affairs, [http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536459/IPOL\\_STU\(2015\)536459\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2015/536459/IPOL_STU(2015)536459_EN.pdf) (Abruf: 2016-02-24).
40. ULD *Positionspapier des ULD zum Safe-Harbor-Urteil des Gerichtshofs der Europäischen Union vom 6. Oktober 2015, C-362/14*. 2015. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, [https://www.datenschutzzentrum.de/uploads/internationales/20151014\\_ULD-Positionspapier-zum-EuGH-Urteil.pdf](https://www.datenschutzzentrum.de/uploads/internationales/20151014_ULD-Positionspapier-zum-EuGH-Urteil.pdf).
41. Kompetenzzentrum-Trusted-Cloud *Thesenpapier – Schweigepflicht bei der Auslagerung von IT-Dienstleistungen*. 2015. Kompetenzzentrum Trusted Cloud, [http://www.trusted-cloud.de/media/content/150129\\_Thesenpapier\\_Schweigepflicht\\_gesamt\\_RZ\\_Ansicht\\_EZ.pdf](http://www.trusted-cloud.de/media/content/150129_Thesenpapier_Schweigepflicht_gesamt_RZ_Ansicht_EZ.pdf) (Abruf: 2015-04-19).
42. Schneider, U.K., *Sekundärnutzung klinischer Daten - Rechtliche Rahmenbedingungen*. 2015, Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin.
43. Weichert, T., *Die Europäische Datenschutz-Grundverordnung – ein Überblick*. DANA, 2016. **2016**(2): S. 48-56.
44. Slawik, M., Zickau, S., Thatmann, D., Repschläger, J., Ermakova, T., Küpper, A., Zarnekow, R., *Innovative Architektur für sicheres Cloud Computing: Beispiel eines Cloud-Ecosystems im Gesundheitswesen*. Proceedings of the 42th Annual Conference of the Gesellschaft für Informatik e.V., 2012: S. 1075-1082.
45. DFG *Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Stellungnahme der Kommission für IT-Infrastruktur für 2016–2020*. 2016. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/foerderung/programme/wgi/kfr\\_stellungnahme\\_2016\\_2020.pdf](http://www.dfg.de/download/pdf/foerderung/programme/wgi/kfr_stellungnahme_2016_2020.pdf) (Abruf: 2016-11-23).
46. EC *Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud*. 2016. European Commission, [http://ec.europa.eu/research/openscience/pdf/realising\\_the\\_european\\_open\\_science\\_cloud\\_2016.pdf](http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf) (Abruf: 2016-12-01).
47. HGF *Die Ressource Information besser nutzbar machen! Positionspapier zum Umgang mit Forschungsdaten in der Helmholtz-Gemeinschaft*. 2016. Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V., <https://www.helmholtz.de/os-positionspapier/> (Abruf: 2016-10-26).
48. Brors, B., Eberhardt, W., Eils, R. et al. *White paper: The Applied & Translational Genomics Cloud (ATGC)*. 2015, <http://www.applied-translational-genomics-cloud.de/joomla/index.php/de/download> (Abruf: 2016-01-25).

49. Diedrich, O., *Container: Apps für Server. Wie Docker-Container die IT industrialisieren.* c't, 2016. **2016**(5): S. 108-110.
50. Seidel, U., *Build, Ship, Run. Das Docker-Universum im dritten Lebensjahr.* ix, 2016. **2016**(4): S. 46-39.
51. Löbe, M., Ganslandt, T., Lotzmann, L. et al., *Simplified Deployment of Health Informatics Applications by Providing Docker Images.* Stud Health Technol Inform, 2016. **228**: S. 643-647.
52. ENISA *Cloud Computing. Benefits, risks and recommendations for information security.* 2009. European Union Agency for Network and Information Security, [https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment/at\\_download/fullReport](https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment/at_download/fullReport) (Abruf: 2016-10-25).
53. BSI *Anforderungskatalog Cloud Computing. Kriterien zur Beurteilung der Informationssicherheit von Cloud-Diensten.* 2016. Bundesamt für Sicherheit in der Informationstechnik, Version 1.0, <https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/CloudComputing/Anforderungskatalog/Anforderungskatalog.pdf> (Abruf: 2016-03-21).



# Strategische Aktivitäten und Fördermaßnahmen im Bereich der IT-Infrastrukturen

*Annette Pollex-Krüger, Johannes Janssen*

## Strategische Aktivitäten

### **Bestandsaufnahme zu Informationsinfrastrukturen – Empfehlungen des Wissenschaftsrats**

Leistungsfähige Infrastrukturen sind ein zentraler Erfolgsfaktor für jede Form wissenschaftlicher Forschung und somit ein wesentlicher Bestandteil des Wissenschaftssystems. Hierzu haben verschiedene Gremien, wie beispielsweise der Wissenschaftsrat (WR) oder die Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK), in den letzten Jahren eine Reihe von Empfehlungen und Konzepten erarbeitet [1-3]. Aus diesen und anderen Stellungnahmen wird deutlich, dass sich die wissenschaftlichen Informationsinfrastrukturen zurzeit in einer höchst dynamischen Entwicklung befinden, die sich aus dem stetigen Wandel des Wissenschaftssystems und seinen zunehmenden Anforderungen insbesondere an die IT-Infrastrukturen speist. Auch die fortschreitende Digitalisierung erzeugt einen zunehmenden Bedarf an Informationsinfrastrukturen, insbesondere im Bereich der medizinischen Forschung.

Bereits 2006 legte die Arbeitsgruppe „Zukunft der Fachinformation“ der damaligen Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK) ihren Abschlussbericht zur Neuausrichtung der öffentlich geförderten Informationseinrichtungen vor [4]. Am Beispiel der Informationseinrichtungen der Leibniz-Gemeinschaft befasst sich der Bericht mit der Neuausrichtung der durch die Digitalisierung und weltweiten Vernetzung veränderten wissenschaftlichen Informationsinfrastruktur. Die Empfehlungen des Berichts beziehen sich auf die Handlungsfelder Informationsdienstleistungen, Lizenzierung, Open Access und Archivierung. Als Kernproblem wurde schon damals die fehlende übergreifende strategische Steuerung der Aktivitäten erkannt. Heterogene Entscheidungsstrukturen führen demzufolge zu Reibungsverlusten und letztlich dazu, dass die Potentiale der bestehenden Infrastrukturen nicht ausgeschöpft werden können.

In seiner umfassenden Bestandsaufnahme „Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020“ formuliert der Wissenschaftsrat [2], unter Berücksichtigung einer eigenen früheren Stellungnahme [1] und eines parallel im Auftrag der GWK erarbeiteten Gesamtkonzepts [3], nicht nur Empfehlungen in Bezug auf die inhaltliche Weiterentwicklung der Informationsinfrastrukturen, sondern bezieht ebenso deren Finanzierung und Planung, ihre Organisation, ihre Nutzbarkeit und tatsächliche Nutzung sowie weitere Rahmenbedingungen mit ein. So betont der Wissenschaftsrat in Bezug auf die Finanzierung von Informationsinfrastrukturen, dass die „Gewährleistung der Verfügbarkeit von Informationsinfrastrukturen als unverzichtbarem Bestandteil des Wissenschaftssystems [...] eine öffentliche Aufgabe“ ist, und dass „Bund und Länder [...] aufgerufen [sind], eine ausreichende Grundfinanzierung der von ihnen getragenen öffentlichen Informationsinfrastruktureinrichtungen, insbesondere auch an Hochschulen, sicher zu stellen“ [2]. Auch wenn Forschungsinfrastrukturen teilweise im Rahmen von Projekten entwickelt werden, sind sie als Infrastrukturen – anders als die Forschung, die sie ermöglichen – grundsätzlich nicht projektförmig. Vielmehr sind sie langfristig bis dauerhaft angelegt und bedürfen daher einer gesicherten und ausreichenden institutionellen Grundfinanzierung [2]. In dem Zusammenhang werden die Förderorganisationen aufgerufen, ihre Vergabekriterien zu ändern und neue Förderprogramme zur verbesserten Finanzierung informationsinfrastrukturbezogener Forschung zu entwickeln. Die DFG hat in einer Stellungnahme zu den Empfehlungen des Wissenschaftsrats diese ausdrücklich begrüßt und sieht hinsichtlich der Koordination der Informationsinfrastrukturentwicklung in Übereinstimmung mit dem Wissenschaftsrat eine besondere Herausforderung darin, dass den Selbstorganisationskräften der wissenschaftlichen Communities bei der Ausgestaltung einer innovativen Informationsinfrastruktur eine besondere Bedeutung zukommen muss. Gleichzeitig bestehe aber auch ein erheblicher Koordinierungsbedarf, um Mehrfacharbeit, Mehrfachfinanzierung und die Inkompatibilität verschiedener Systeme zu vermeiden.

Weiterhin weist der Wissenschaftsrat darauf hin, dass Investitionsentscheidungen für konkrete Forschungsinfrastrukturen stets zu Pfadabhängigkeiten führen, die intendierte wie nicht intendierte Folgen insbesondere für das Verhältnis zwischen universitärer und außeruniversitärer Forschung, zwischen den Disziplinen sowie für innerdisziplinäre Entwicklungen zeitigen [1]. So spricht der Wissenschaftsrat die Empfehlung aus, dass „bei der Ansiedelung von Informationsinfrastrukturen [...] alle Sektoren des Wissenschaftssystems berücksichtigt werden [sollten]“. Insbesondere seien „die Hochschulen in größerem Umfang als Träger von Informationsinfrastrukturen zu berücksichtigen, um auf diese Weise sicher zu stellen, dass es in allen Teilbereichen des Wissenschaftssystems Informationsinfrastrukturen gibt“ [1]. Hintergrund dieser Empfehlung ist die

Annahme, dass sich Forschungsmöglichkeiten und Forschungsbereiche im Umfeld von Forschungsinfrastrukturen konzentrieren können. Eine weitgehende Entkopplung solcher Forschungsbereiche von der hochschulischen Forschung, mit unerwünschten Folgen beispielsweise hinsichtlich der Verbindung von Forschung und Lehre und somit auch der Gewinnung von wissenschaftlichem Nachwuchs, soll vermieden werden [1]. Weitere Empfehlungen beziehen sich auf eine ausreichende Ausstattung von Informationsinfrastrukturen mit qualifiziertem Personal; überdies sollten infrastrukturbezogene wissenschaftliche Leistungen in Qualifikationsverfahren ergänzend anerkannt werden.

Der Wissenschaftsrat schlägt in diesem Zusammenhang ein neues, erweitertes Verständnis von Forschungsinfrastrukturen vor, welche sich lange Zeit auf die besonders investitionsintensiven Großgeräte für die naturwissenschaftliche Grundlagenforschung beschränkten. Damit werden europäische Diskussionen zum Forschungsinfrastrukturbegriff aufgenommen, wie sie etwa im Kontext von ESFRI (s.u.) oder MO Forum on Research Infrastructures (s.u.) entwickelt wurden. Ausgehend von dem unter [1] beschriebenen, erweiterten Begriff der Forschungsinfrastrukturen unterscheidet der Wissenschaftsrat vier Kategorien:

1. Großgeräte, sowie Forschungsplattformen wie z. B. Forschungsschiffe, u. ä.,
2. informationstechnische und e-Infrastrukturen,
3. soziale Infrastrukturen und
4. Informationsinfrastrukturen.

Unter sozialen Infrastrukturen werden Einrichtungen verstanden, die dem (inter)disziplinären Austausch von aktuellen und der Entwicklung von neuen Forschungsfragen dienen. Grundsätzlich gibt es zwischen den Forschungsinfrastrukturen zahlreiche Mischformen und Übergänge. So weisen beispielsweise viele Informationsinfrastrukturen auch Charakteristika einer sozialen Infrastruktur auf, d. h. sie dienen auch der wissenschaftlichen Kommunikation. Grundsätzlich kommt den Informationsinfrastrukturen nach Ansicht des WR aufgrund ihrer Disziplin-übergreifenden Relevanz eine zentrale Bedeutung zu.

Mit seinen im Oktober 2016 verabschiedeten Empfehlungen „Perspektiven der Universitätsmedizin“ [5] schließt der Wissenschaftsrat an frühere Strukturempfehlungen zur Universitätsmedizin an. Die aktuelle Stellungnahme verfolgt das Ziel, die Situation der universitären Medizin vor dem Hintergrund der aktuellen gesellschaftlichen und wissenschaftlichen Herausforderungen zu analysieren und Empfehlungen zu ihrer weiteren Entwicklung an der Schnittstelle des Wissenschafts- und des Gesundheitssystems zu erarbeiten. Als eines der zentralen Probleme der Universitätsmedizin wird

formuliert, dass die derzeitige Ausstattung mit Forschungs- und informationstechnologischen Infrastrukturen den wachsenden Aufgaben nicht hinreichend gerecht wird und einer substanziellen Verbesserung bedarf. Unter anderem führten die Ressourcendefizite zu einer Dominanz der Versorgungsaufgaben, welche die Freiräume für akademische Kernaufgaben einschränken und zu einer fortschreitenden Marginalisierung der akademischen Zielsetzung beitragen. Bei der Weiterentwicklung komplexer Forschungsinfrastrukturen kommt es nach Ansicht des Wissenschaftsrates vor allem auf den Aufbau zentraler Strukturen und deren Verzahnung mit lokalen Forschungsinfrastrukturen an. Im Bereich der IT-Infrastrukturen bewertet der Wissenschaftsrat das Förderkonzept Medizininformatik des BMBF als wichtigen Impulsgeber für den Aufbau solcher Strukturen. Er empfiehlt, die Förderinitiative der Wissenschaftsseite durch komplementäre Aktivitäten der Gesundheitsseite zu flankieren, um eine nachhaltige Wirkung zu erreichen.

Angesichts der Investitionsbedarfe der Hochschulen im Allgemeinen wie der Universitätsmedizin im Besonderen erneuert der Wissenschaftsrat seine an Bund und Länder gerichtete Bitte, Möglichkeiten einer langfristigen gemeinsamen Finanzierung von Infrastrukturen an Hochschulen zu prüfen.

### **Rat für Informationsinfrastrukturen**

Auf Empfehlung des Wissenschaftsrats [2] hat die Gemeinsame Wissenschaftskonferenz des Bundes und der Länder (GWK) ab Herbst 2014, zunächst befristet auf vier Jahre, den Rat für Informationsinfrastrukturen (RfII) eingerichtet.<sup>1</sup> In seiner Auftakterklärung vom Juni 2015 formuliert der Rat als Gegenstand seiner Tätigkeit die Gestaltung zukunftsfähiger Informationsinfrastrukturen für die Wissenschaft.<sup>2</sup> Eine zentrale Aufgabe des RfII ist es, die Entwicklung der wissenschaftlichen Informationsinfrastrukturen in ihren vielfältigen Aktivitäten besser aufeinander abzustimmen und miteinander zu verzahnen. Weiterhin soll sich der Rat auf Systemebene den strategischen Zukunftsfragen dieses Wissenschaftsbereichs widmen, die Selbstorganisationsprozesse in der Wissenschaft stärken, Kooperationsmöglichkeiten ausloten, sowie neue Handlungsfelder identifizieren und wettbewerbliche Verfahren anstoßen. Mit der Gründung des Rats für Informationsinfrastrukturen folgt die GWK der Einschätzung des Wissenschaftsrats, dass für eine strategische Weiterentwicklung des Gesamtsystems der Informationsinfrastrukturen in Deutschland ein übergeordnetes Koordinierungs- und Beratungsgremium erforderlich ist. Dieses soll im Auftrag von Bund und Ländern, aber auch in eigener

1 Siehe <http://www.gwk-bonn.de/uebergreifende-wissenschafts-und-forschungspolitische-themen/informationsinfrastruktur/> (06.11.2016).

2 Siehe <http://www.rfii.de/de/category/dokumente/> (06.11.2016).

Initiative tätig werden. Die Mitglieder des Rats für Informationsinfrastrukturen werden von der GWK berufen und setzen sich zusammen aus Vertretern von Informationsinfrastruktureinrichtungen und der Wissenschaft sowie Vertretern von Bund, Ländern und des öffentlichen Lebens. In der aktuellen Arbeitsperiode setzt sich der RfII mit dem Themenkomplex „Forschungsdaten – Nachhaltigkeit – Internationalität“ auseinander.

Im Juni 2016 hat der Rat mit seinem Positionspapier „Leistung aus Vielfalt“ umfassende Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland vorgelegt [6]. In seiner Bestandsaufnahme benennt der RfII die Handlungsbedarfe. Trotz einiger guter Beispiele sei die Landschaft in Deutschland bislang noch durch überwiegend schwach koordinierte, projektförmige Initiativen geprägt. Eine Grundversorgung der Forschenden mit niedrigschwelligen Services für das Forschungsdatenmanagement fehlt, es bleibt derzeit bei Anstrengungen einzelner Institutionen und Organisationen. Die Empfehlungen des RfII lassen sich in Kürze wie folgt zusammenfassen:

- ▶ Fördermechanismen nachhaltig gestalten (Verstetigung von Infrastrukturprojekten),
- ▶ mit dem Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI) die Effizienz und Koordination durch eine (verteilte) nationale Infrastruktur sicherstellen,
- ▶ Datenverantwortung und gute Datenkultur befördern,
- ▶ übergreifendes Monitoring und Qualitätssicherung einrichten und
- ▶ Personalentwicklung auf allen Ebenen betreiben.

Der RfII geht davon aus, dass zur Sicherung der Leistungsfähigkeit der deutschen Wissenschaft erheblicher Investitions- und Finanzierungsbedarf auf verschiedenen Ebenen des Gesamtsystems besteht. Das Positionspapier „Leistung aus Vielfalt“ enthält neben den Empfehlungen einen ausführlichen Anhang mit Daten und Fakten, darunter 25 Kurztexte zur Klärung von Begriffen rund um das Thema Forschungsdaten.

In der Folge wird der Rat seine kürzlich veröffentlichten Empfehlungen zu einer Nationalen Forschungsdateninfrastruktur (NFDI) weiter ausgestalten und den begonnen Dialog mit der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) über die Errichtung einer NFDI fortführen. Ein Schwerpunkt der weiteren Arbeiten wird auf der Analyse nationaler wie internationaler Entwicklungen im Bereich Forschungsdaten liegen und dabei auch die Anschlussfähigkeit der deutschen Aktivitäten an europäische und globale Entwicklungen und Initiativen wie der European Science Cloud [7] in den Blick nehmen. Darüber hinaus wird er sich mit Rechtsfragen rund um Datenschutz und Datensicherheit befassen.

## Forum Gesundheitsforschung

Am 16.11.2015 hat sich auf Einladung des BMBF als weiteres Beratungsgremium das Forum Gesundheitsforschung<sup>3</sup> konstituiert. Seine Mitglieder sind Vertreter des Wissenschaftsrats, der DFG, der vier außeruniversitären Forschungsorganisationen Helmholtz-Gemeinschaft, Max-Planck-Gesellschaft, Leibniz-Gemeinschaft und Fraunhofer Gesellschaft, sowie des Medizinischen Fakultätentages (MFT), des Verbandes der Universitätsklinika Deutschland (VUD), der Deutschen Zentren der Gesundheitsforschung (DZG) sowie der Wirtschaft. Insbesondere soll das Forum Strategien zur Weiterentwicklung der Gesundheitsforschung diskutieren und Vorschläge für drängende Forschungsfragen sowie forschungspolitische Maßnahmen erarbeiten. Als erste Arbeitsgruppe wurde die AG „Infrastrukturen in den Lebenswissenschaften“ eingerichtet. Diese soll auf der Basis einer Bestandsaufnahme und Bedarfsanalyse einen Überblick über den möglichen strategischen Handlungsbedarf für den zukunftsfähigen Aus- und Aufbau von Geräte-Infrastrukturen in der biomedizinischen Forschung erarbeiten. Das Themenfeld „Next Generation Sequencing“ (NGS) – aktuell ebenfalls bei der DFG intensiv diskutiert – steht beispielhaft für die Entwicklung einer verteilten Forschungsinfrastruktur, bei der es gilt, vorhandene Ressourcen und wachsende Bedarfe in angemessener Weise zu reflektieren und eine für die Wissenschaft adäquate Nutzung zu ermöglichen. Dabei rücken klassische Investitionen in der Gesamtkostenbetrachtung zunehmend in den Hintergrund, während Personal und projektbezogene Sequenzierkosten zu maßgeblichen Faktoren werden. Die zweite Arbeitsgruppe „Wertschöpfungskette“ soll sich mit den Innovationshemmnissen beschäftigen, die insbesondere die schnelle Translation der Forschungsergebnisse in die medizinische Anwendung verhindern, sowie Lösungsansätze und Umsetzungsstrategien erarbeiten.

In seiner Aufgabe als Beratungsgremium für das BMBF folgt das Forum Gesundheitsforschung dem Gesundheitsforschungsrat nach, der im Frühjahr 2013 aufgelöst wurde.

## Nationale Roadmap für Forschungsinfrastrukturen

Für die Einrichtung und langfristige Bereitstellung von Forschungsinfrastrukturen sind zunehmend erhebliche Mittel erforderlich, bei gleichzeitig knapper werdenden Budgets der öffentlichen Haushalte. Um die verfügbaren Mittel möglichst effizient einzusetzen und die Realisierung neuer Forschungsinfrastrukturen von nationaler und internationaler Bedeutung langfristig strategisch planen zu können, hat das BMBF von 2011-2013 eine Pilotphase für die Erstellung einer Roadmap für Forschungsinfrastrukturen

---

3 Siehe <http://www.gesundheitsforschung-bmbf.de/de/forum.php> (06.11.2016).

durchgeführt, die im Jahr 2016 als Roadmap-Prozess fortgeführt wird. Die Priorisierung der Forschungsinfrastrukturen erfolgt auf Basis eines einheitlich strukturierten Verfahrens, wobei die wissenschaftliche Bewertung vom Wissenschaftsrat entwickelt und durchgeführt wird. Die wirtschaftliche Bewertung der Konzepte (Kostenschätzung) wird vom Projektträger im DLR durchgeführt. Das BMBF entscheidet dann unter Berücksichtigung der gesellschaftlichen Relevanz und der finanziellen Realisierbarkeit der geplanten Forschungsinfrastrukturen über die Aufnahme in die Roadmap. Mit der Aufnahme ist gleichzeitig eine grundsätzliche Förderabsicht des BMBF verbunden. Aufbauend auf den Erkenntnissen aus dem Pilotprozess wurde im August 2015 der „Nationale Roadmap-Prozess Forschungsinfrastrukturen“ öffentlich bekanntgegeben.<sup>4</sup> An dem Prozess können sich alle interessierten Hochschulen und außeruniversitären Forschungseinrichtungen beteiligen, die den Aufbau neuer Forschungsinfrastrukturen planen. Der Prozess fokussiert auf die Investitionskosten und definiert Schwellenwerte (50 Mio. Euro resp. 20 Mio. Euro in den Bereichen Geistes-/Sozialwissenschaften/Bildungsforschung). Anfang 2016 waren Konzepte einzureichen. Entscheidungen sind Anfang 2018 zu erwarten. Die Nationale Roadmap Forschungsinfrastrukturen soll in Abständen aktualisiert werden.

Forschungsinfrastrukturen im Sinne des Roadmap-Prozesses des BMBF sind umfangreiche und langlebige Ressourcen oder Serviceeinrichtungen für die Forschung in allen Wissenschaftsgebieten. Zurzeit befinden sich 27 Projekte auf der Nationalen Roadmap für Forschungsinfrastrukturen, die überwiegend international kofinanziert sind. Im Bereich der bio-medizinischen Forschung sind die folgenden sechs Infrastrukturen Teil der Nationalen Roadmap: ECRIN (European Clinical Research Infrastructures Network), Infrafrontier (Systemische Phänotypisierung, Archivierung und Verteilung von Mausmodellen), IPL (In-Vivo-Pathophysiologie-Labor der Helmholtz-Gemeinschaft), die NAKO Gesundheitsstudie, SHARE (Survey of Health, Ageing and Retirement in Europe) und EU-OPENSREEN (Open screening platforms for chemical biology). Auf Grundlage der Nationalen Roadmap sind Vereinbarungen und Vernetzungen mit nationalen und internationalen Partnern möglich. Hierzu zählt insbesondere das Europäische Strategieforum für Forschungsinfrastrukturen (European Strategy Forum on Research Infrastructures, ESFRI).

## **Forschungsinfrastrukturen und Roadmaps einzelner Forschungsorganisationen**

Die Helmholtz-Gemeinschaft hat im Jahr 2011 eine Roadmap zu Forschungsinfrastrukturen veröffentlicht, die im Jahr 2015 aktualisiert<sup>5</sup> wurde. In dieser sind die wichtigsten

---

4 Siehe <https://www.bmbf.de/foerderungen/bekanntmachung.php?B=1088> (06.11.2016).

5 Siehe [https://www.helmholtz.de/publikationen/Helmholtz\\_Roadmap\\_2015\\_web\\_korr\\_150921.pdf](https://www.helmholtz.de/publikationen/Helmholtz_Roadmap_2015_web_korr_150921.pdf) (06.11.2016).

zukunftsrelevanten Forschungsinfrastrukturen der Helmholtz-Gemeinschaft aufgelistet, die in den kommenden Jahren für die strategische Umsetzung ihres wissenschaftlichen Portfolios relevant sind. Für den Forschungsbereich Gesundheit ist als einzige Forschungsinfrastruktur die bundesweite NAKO Gesundheitsstudie gelistet.

Die Leibniz-Gemeinschaft betreibt eine Reihe wichtiger Forschungsinfrastrukturen, die in einer Broschüre gebündelt dargestellt werden.<sup>6</sup> In einer weiteren Broschüre werden Befunde und Empfehlungen für den Bereich Forschungsinfrastrukturen vorgestellt und auch eine Reihe von offenen Fragen und Impulsen für eine weiterreichende Debatte formuliert.<sup>7</sup> Adressiert werden Fragen der Qualitätsbestimmung von Forschungsinfrastrukturen, des nachhaltigen Betriebs sowie Potenziale und Strategien infrastrukturorientierter Nachwuchsentwicklung.

Die Deutsche Hochschulmedizin e.V. als gemeinsamer Dachverband von VUD und MFT hat ebenfalls eine Roadmap für die Weiterentwicklung der IT-Infrastruktur vorgelegt.<sup>8</sup> In dem Positionspapier „Medizinischer Fortschritt braucht leistungsstarke IT-Lösungen“ werden die besonderen Herausforderungen der Universitätsmedizin zwischen medizinischer Versorgung, akademischer Lehre und international wettbewerbsfähiger Forschung dargestellt, sowie der Weiterentwicklungs- und Handlungsbedarf beschrieben.

Diese Darstellungen belegen, dass Forschungsinfrastrukturen weiter zunehmend in den Blickwinkel der Forschungsorganisationen und Verbände rücken, was auch als Indiz des wachsenden Einflusses der Infrastrukturen auf das Wissenschaftssystem insgesamt gewertet werden kann.

## **ESFRI – European Strategy Forum on Research Infrastructures**

Auf europäischer Ebene hat vor mehr als zehn Jahren das Europäische Strategieforum für Forschungsinfrastrukturen (ESFRI)<sup>9</sup> die Erarbeitung einer europäischen Forschungsinfrastruktur-Roadmap begonnen und damit vergleichbare Prozesse in den einzelnen europäischen Ländern angestoßen (s.o.). Das Forum ESFRI wurde im April 2002 initiiert mit der Aufgabe, einen Beitrag zu einer europäischen Strategie der Forschungsinfrastrukturen zu leisten und damit die wissenschaftliche Integration innerhalb Europas zu befördern. Der erste Report wurde 2006 veröffentlicht mit zunächst 35 Vorhaben. Zwei weitere Überarbeitungen der ESFRI-Roadmap wurden 2008 und 2010 publiziert.

---

6 Siehe <http://www.leibniz-gemeinschaft.de/forschungsinfrastrukturen-in-der-leibniz-gemeinschaft/> (06.11.2016).

7 Siehe <http://www.leibniz-gemeinschaft.de/forschungsinfrastrukturen-im-wissenschaftssystem/> (06.11.2016).

8 Siehe [http://www.uniklinika.de/14-07-09\\_Deutsche\\_Hochschulmedizin\\_e.V.\\_IT-Positionspapier\\_Final.pdf](http://www.uniklinika.de/14-07-09_Deutsche_Hochschulmedizin_e.V._IT-Positionspapier_Final.pdf) (07.11.2016).

9 Siehe <http://ec.europa.eu/research/infrastructures/esfri> (29.09.2016).

Weiterhin wurde mit Unterstützung durch ESFRI von der Europäischen Kommission eine neue Rechtsform für Forschungsinfrastrukturen geschaffen, das „European Research Infrastructure Consortium (ERIC)“. Inzwischen wurden auf europäischer Ebene 10 ERICs gegründet. Nach Vorlage der ESFRI-Roadmap 2010 begleitete ESFRI zunächst den Implementierungsprozess auf europäischer und nationaler Ebene. Die Projekte wurden mit Hilfe der Europäischen Kommission zunächst evaluiert und es wurden Empfehlungen zur weiteren Implementierung formuliert.

Inzwischen liegt als nächste Aktualisierung die ESFRI-Roadmap 2016 vor. Diese enthält 21 Infrastruktur-Projekte, welche in ihrer Entwicklung weit fortgeschritten sind, sowie 29 sogenannte ESFRI Landmarks. So werden in der Roadmap 2016 Forschungsinfrastrukturen bezeichnet, die ihre Implementierungsphase Ende 2015 erreicht haben. Zu den ESFRI Landmarks gehören die Projekte BBMRI-ERIC, EATRIS-ERIC, ECRIN-ERIC, ELIXIR und INFRAFRONTIER. Im Bereich der biomedizinischen Forschungsinfrastrukturen wird insbesondere ELIXIR (The European Life-Science Infrastructure for Biological Information) als für Europa strategisch relevant priorisiert, da durch diese verteilte Infrastruktur große Mengen an lebenswissenschaftlichen Daten für die Forschung verfügbar gemacht werden sollen. Aber auch BBMRI, INFRAFRONTIER, EATRIS und ECRIN empfiehlt das Europäische Strategieforum Mitgliedsländern zur Implementierung bzw. weiteren Unterstützung und Vernetzung.<sup>10</sup>

### **European Science Foundation (ESF) – MO Forum on Research Infrastructures**

Europäische Forschungsorganisationen und Forschungsförderer haben sich in der European Science Foundation (ESF)<sup>11</sup> mit Fragen der Finanzierung, Organisation und des Managements von europäischen Forschungsinfrastrukturen beschäftigt und von 2010 bis 2013 das „ESF Member Organisation Forum on Research Infrastructures“ in Leben gerufen. Unter anderem wurde, mit Unterstützung der Europäischen Kommission, das MERIL<sup>12</sup> Projekt (Mapping of the European Research Infrastructure Landscape) durchgeführt, welches eine umfassende Bestandsaufnahme von Forschungsinfrastrukturen in Europa zum Ziel hat. Während ESFRI sich mit der Planung und dem Aufbau von neuen Forschungsinfrastrukturen beschäftigt, fokussierte sich das MO Forum on Research Infrastructures auf bestehende Infrastrukturen. So wurden beispielsweise Qualitätsstandards in Bezug auf die Organisation, das Management und besonders den Zugang zu Forschungsinfrastrukturen (provision of access) definiert, die u. a. Eingang

---

10 Siehe [https://ec.europa.eu/research/infrastructures/pdf/ESFRI\\_projects\\_for\\_impl\\_7\\_april\\_2014.pdf](https://ec.europa.eu/research/infrastructures/pdf/ESFRI_projects_for_impl_7_april_2014.pdf) (29.09.2016).

11 Siehe <http://www.esf.org/home.html> (05.10.2016).

12 Siehe <http://portal.meril.eu/> (05.10.2016).

in die European Charter for Access to Research Infrastructures<sup>13</sup> fanden. In einer Kooperation zwischen MERIL und ESFRI wurden sogenannte „Common Features“ von Forschungsinfrastrukturen festgelegt.<sup>14</sup> Mit Beginn des Jahres 2016 wurde die ESF in eine Service-basierte Organisation umgewandelt. Die ursprünglichen forschungsunterstützenden Aktivitäten werden abgebaut und evaluierungsunterstützende Services entwickelt. MERIL wird weiterhin federführend von der ESF (mit Unterstützung durch EKT<sup>15</sup> und APRE<sup>16</sup>) in einem von der Europäischen Kommission finanzierten Projekt weiterentwickelt. Ein Upgrade beinhaltet Anpassungen am Datenmodell wie an der Portalfunktionalität und schließt Nutzungserfahrungen ein, mit dem Ziel, Entscheidungsträgern und Nutzern von Forschungsinfrastrukturen umfassende Informationen bereitzustellen.

### **Science Europe – Working Group on Research Infrastructures**

Über die 2011 ins Leben gerufene Organisation Science Europe gestalten Europäische Forschungsorganisationen und Forschungsförderer transnationale Themen und bringen ihre Perspektive in die europäischen Debatten dazu ein. Die im MO Forum begonnenen Aktivitäten zu Forschungsinfrastrukturen werden seit 2013 in diesem Sinne unter dem Dach von Science Europe fortgeführt. Science Europe hat hierzu eine Arbeitsgruppe<sup>17</sup> etabliert, in der das gemeinsame Verständnis der nationalen Organisationen zu Forschungsinfrastrukturen weiterentwickelt wird und spezifische Workshops veranstaltet werden können.

### **Expertenkommission Forschung und Innovation (EFI)**

Neben dem Wissenschaftsrat leistet auch die Expertenkommission Forschung und Innovation (EFI) wissenschaftliche Politikberatung für die Bundesregierung und legt regelmäßig Gutachten zur Forschung, Innovation und technologischen Leistungsfähigkeit Deutschlands vor. Im EFI-Jahresgutachten 2014<sup>18</sup> war die Forschung und Innovation in der Hochschulmedizin eines der Kernthemen. Dort wird, allerdings nur übersichtsartig, auf die Nutzung von Informations- und Kommunikationstechnologien (IKT) in der medizinischen Forschung eingegangen. Die zunehmende Etablierung der Open Source Lösungen i2b2 und SHRINE in Deutschland durch die Hochschulklinika Erlangen-Nürnberg, Göttingen und Leipzig und die Rolle der TMF werden kurz er-

---

13 Siehe [https://ec.europa.eu/research/infrastructures/pdf/2015\\_charterforaccessto-ris.pdf](https://ec.europa.eu/research/infrastructures/pdf/2015_charterforaccessto-ris.pdf) (05.10.2016).

14 Siehe [http://www.esf.org/fileadmin/Public\\_documents/Publications/mof\\_research\\_infrastructures.pdf](http://www.esf.org/fileadmin/Public_documents/Publications/mof_research_infrastructures.pdf) (05.10.2016).

15 Greek National Documentation Centre ([www.ekt.gr/en](http://www.ekt.gr/en)) (24.10.2016)

16 Agency for the Promotion of European Research ([www.apre.it/en](http://www.apre.it/en)) (24.10.2016)

17 Siehe <http://www.scienceeurope.org/policy/working-groups/Research-Infrastructures> (05.10.2016).

18 Siehe [http://www.e-fi.de/fileadmin/Gutachten\\_2014/EFI\\_Gutachten\\_2014.pdf](http://www.e-fi.de/fileadmin/Gutachten_2014/EFI_Gutachten_2014.pdf) (05.10.2016).

wähnt. Die Expertenkommission konstatiert mit Besorgnis die deutsche IKT-Schwäche auch im Bereich der Medizinforschung und fordert von Bund und Ländern einen Aktionsplan zur Nutzung der Datenmengen in der medizinischen Forschung. Dieser Aktionsplan sollte nach Ansicht der Expertenkommission in die Digitale Agenda integriert werden. In den EFI-Jahresgutachten 2015 und 2016 waren die Medizin bzw. die Gesundheitsforschung keine Kernthemen, so dass diese hier nicht diskutiert werden.

### **Empfehlungen der Kommission für IT-Infrastruktur der DFG**

Die Kommission für IT-Infrastruktur der DFG (KfR, bis 2007 DFG-Kommission für Rechenanlagen) veröffentlicht regelmäßig Empfehlungen für Planung und Betrieb der IT-Infrastruktur an Hochschulen und Universitätsklinika, zuletzt Anfang 2016 für den Zeitraum von 2016 – 2020 („Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Stellungnahme der Kommission für IT-Infrastruktur für 2016-2020“) [8]. In diesem Papier schreibt sie nicht einfach ältere Empfehlungen fort, sondern spricht mit seiner neuen Strukturierung des Inhalts bewusst die verschiedenen Ebenen an, in welchen IT-Ausstattungsaspekte an Hochschulen adressiert werden müssen (Ministerien, Hochschulleitungen, Rechenzentren, Wissenschaft). So betont sie die Notwendigkeit der Einbindung von IT-Konzepten in klare Strategien und Schwerpunktsetzungen der Hochschulen. Insbesondere empfiehlt sie den Aufbau standortübergreifender Lösungen. Die Empfehlungen sollen die konkrete Arbeit an IT-Konzepten und daraus abgeleiteten Infrastruktur-Anträgen unterstützen.

In dem Kapitel zur Universitätsmedizin wird ausgeführt, dass hier die enge Verknüpfung von Forschung und Lehre mit der Krankenversorgung auch die IT-Systeme durchdringt und spezielle Herausforderungen mit sich bringt, etwa in Bezug auf patientenbezogene Daten, Forschungsdaten und Translation. Die Informationstechnologie betrifft zunehmend alle technisierbaren Prozesse der Diagnostik und Therapie sowie der Logistik, bis hin zu einer neuen Generation von Entscheidungsunterstützungs- bzw. Wissensaufbereitungssystemen, die wiederum großen Einfluss auf die Ausbildung zur Ärztin bzw. zum Arzt haben werden. Erheblicher neuer Integrationsbedarf wird bei der Einbindung neuer Datenarten und großer Datenmengen aus den Omics-Laboren gesehen sowie aus verschiedensten Anwendungen der Sensortechnik, der Bilddiagnostik und Laborverfahren. Letztere – einschließlich der oft angekoppelten Biobanksysteme – verlangen hoch-industrialisierte Arbeits- und Datenverarbeitungsprozesse, um nach standardisierten Regeln qualitätsgesicherte Daten liefern zu können. Die Dokumentation der Qualität der eingesetzten Verfahren über Metadaten werde – wie überall in der Forschung und Versorgung – eine größere Rolle als bisher bekommen. Durch das E-Health-Gesetz sei absehbar, dass in den kommenden Jahren alle Universitätskliniken

ihre Zusammenarbeit mit anderen Versorgungseinrichtungen in Deutschland ausbauen und ihre vorhandenen IT-Installationen anpassen müssen. Dabei kristallisiert sich heraus, dass die einrichtungsübergreifenden elektronischen Patientenakten (ePA) eine wesentliche Bedeutung für die Bündelung isolierter Verfahren der Gesundheitstelematik besitzen und die Basis einer Gesundheitstelematikinfrastruktur bilden. IT-Infrastrukturen zur überregionalen Integration von Forschungsdaten werden bis 2020 ebenfalls eine strategische Bedeutung erlangen. Weitere Herausforderungen sind die Systemintegration von Medizinprodukten sowie der Einsatz terminologischer Ordnungssysteme, wie beispielsweise die international etablierte Nomenklatur SNOMED-CT oder das terminologische System LOINC, und internationale Interoperabilitätsstandards, von denen die IHE-Initiative am vielversprechendsten erscheint. Diese bisher in Deutschland vernachlässigten Komponenten bzw. Standards werden nach Ansicht der KfR konsequenter genutzt werden müssen.

Die zunächst als Addendum zu früheren Empfehlungen für die Jahre 2011-2015 herausgegebene Stellungnahme zur Nutzung von Cloud-Diensten [9] ist in die aktuelle Stellungnahme für die Jahre 2016-2020 [8] integriert worden. Neben den Vorteilen für den Nutzer, als die insbesondere geringere Kosten und bessere Skalierbarkeit genannt werden, diskutiert die KfR auch Risiken in Bezug auf Sicherheit, Datenschutz und Nachhaltigkeit. Diese und weitere Aspekte werden differenziert betrachtet und Empfehlungen bei der Inanspruchnahme von Cloud-Diensten ausgesprochen.<sup>19</sup>

### **DFG-Positionspapier „Die digitale Transformation weiter gestalten“**

Mit dem 2012 vorgelegten Positionspapier „Die digitale Transformation weiter gestalten – Der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung“ [10] hat der Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme der DFG seine Analysen und Vorschläge zur Neuausrichtung der Informationsinfrastrukturen in Deutschland von 2006 weiterentwickelt. Insbesondere empfiehlt der Ausschuss, Förderaktivitäten im Bereich der Informationsinfrastrukturen für Forschungsdaten weiterzuführen und durch die Einrichtung eines Förderprogramms zu verstetigen – was die DFG mit der Einrichtung des Programms „Informationsinfrastrukturen für Forschungsdaten“ 2013 umsetzte. Zusammengefasst kommt in dem Positionspapier zum Ausdruck, dass die strategische Weiterentwicklung aller für Informationsinfrastrukturen zentralen Handlungsfelder als grundlegende Aufgabe der DFG als einer Selbstverwaltungsorganisation der Wissenschaft angesehen wird.

---

<sup>19</sup> Siehe hierzu auch das Kapitel zum Cloud-Computing im vorliegenden Report.

## **Stellungnahmen und Empfehlungen der DFG-Senatskommission für Grundsatzfragen in der Klinischen Forschung**

Für den Bereich der Klinischen Forschung erarbeitet die Ständige Senatskommission für Grundsatzfragen in der Klinischen Forschung der DFG regelmäßig verschiedene Denkschriften, Empfehlungen und Stellungnahmen. Besonders hervorzuheben in Bezug auf IT-Infrastrukturen sind die „Stellungnahme zu den strukturellen Rahmenbedingungen der klinischen Forschung in der deutschen Universitätsmedizin“ [11] von 2014 und die 2015 erschienenen „Empfehlungen zur Weiterentwicklung der Klinischen Forschung an der deutschen Universitätsmedizin in den Jahren 2015 – 2025“ [12]. Es ist absehbar, so die Analyse, dass sich in den kommenden zehn Jahren die Transformation der Medizin von einem in der Vergangenheit primär qualitativen in ein zunehmend quantitatives Forschungsgebiet beschleunigt fortsetzen wird. Um die Leistungsfähigkeit der Universitätsmedizin auf wichtigen Gebieten zu sichern, werden in absehbarer Zeit erhebliche Investitionen für langfristig gesicherte Personal- und Infrastrukturen an den und für die Universitätsklinika erforderlich sein, zusammen mit einer Neuregelung der Finanzierungsmodalitäten zwischen Bund und Ländern.

Als besondere Herausforderung für die nächsten zehn Jahre wird u. a. der Bereich der Infrastrukturen benannt [12]. Hier erfordern beispielsweise die individualisierte Medizin sowie die Durchführung von klinischen Studien den umfassenden und kompetenten Einsatz von medizinischer Informatik, Statistik, Bildgebung und technologischen Infrastrukturen wie Bio- und Datenbanken. Ebenso muss ein hohes Maß an Standardisierung und Vereinheitlichung erreicht werden – sowohl hinsichtlich der erhobenen Parameter, der Dokumentation, der Formate, als auch der Verfügbarmachung und Archivierung der Daten. Ebenso verlangt diese Form der Medizin in der Regel eine multilaterale, überregionale oder sogar internationale Vernetzung, mit entsprechend hohen Anforderungen an das Wissenschaftsmanagement, die erforderlichen Ressourcen und Kompetenzen sowie die übergreifende Abstimmung der nötigen Standards.

Mit dem Ziel, standortübergreifende Standards und Rahmenbedingungen zu entwickeln, wird eine zentrale und integrative Technologie- und Methodenplattform für die Daten- und Informationstechnologie (Nationale Plattform für Daten- und Informationstechnologie) mit permanenter Geschäftsstelle als gemeinsame, ständige Einrichtung der deutschen Universitätsklinika benötigt [12]. Die Plattform sollte unter anderem die Geschäftsbereiche „Beratungsstellen für technische, rechtliche und ethische Angelegenheiten“, „Entwicklung von Software-Lösungen“, „Vernetzung und Standardisierung“ sowie entsprechende Diskussionsforen umfassen. Eine solche Einrichtung erfordert eine langfristig angelegte, gemeinsame Finanzierung durch Bund und Länder.

Die Ähnlichkeit des vorgeschlagenen Konstrukts mit der bestehenden Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF) ist frappierend. Vorarbeiten der TMF werden in dem Bericht entsprechend gewürdigt. Aus ökonomischen und inhaltlichen Gründen sollte eine Parallelentwicklung vermieden und eher eine abgestimmte Weiterentwicklung bestehender Strukturen angestrebt werden.

### **DFG-Portal für Forschungsinfrastrukturen – RIsources**

Mit dem Internetportal „RIsources“<sup>20</sup> hat die DFG ein Informationsportal zu wissenschaftlichen Forschungsinfrastrukturen geschaffen, das sämtliche Einrichtungen oder Plattformen erschließen soll, die Ressourcen und Dienstleistungen für Forschungsvorhaben bereitstellen. Der Begriff „Forschungsinfrastruktur“ reicht in diesem Kontext von dezidierten Geräteangeboten und Analysedienstleistungen bis hin zu Informationsinfrastrukturen wie Bibliotheken, Daten- und Open Access Repositorien oder virtuellen Forschungsumgebungen. Auf diese Weise soll dazu beigetragen werden, vorhandene Ressourcen effektiver zu nutzen, externen Nutzern einen Zugang zu benötigten Technologien beziehungsweise Informationsquellen zu ermöglichen und dabei eine professionelle Unterstützung durch die Einrichtungen anzubieten. Forschungsinfrastrukturen, die in das Portal aufgenommen werden möchten, können sich bei der DFG registrieren. Die Bedingungen für eine Registrierung finden sich in den DFG-Leitlinien zu Anforderungen an Forschungsinfrastrukturen, die sich an den „Basic Requirements for Research Infrastructures in Europe“ der ESF orientieren.<sup>21</sup>

### **MFT-TMF Symposium „Infrastrukturen und Rahmenbedingungen für die medizinische Forschung“**

Der Medizinische Fakultätentag (MFT) und die TMF haben 2012 ein gemeinsames Symposium zum Thema „Infrastrukturen und Rahmenbedingungen für die medizinische Forschung“ durchgeführt [13]. Auf diesem wurden der Aufbau geeigneter Infrastrukturen und die gemeinsame Gestaltung der Rahmenbedingungen als wesentliche Grundlage der patientenorientierten medizinischen Forschung diskutiert. MFT und TMF wiesen darauf hin, dass die Anforderungen an Forschungsvorhaben, an Daten- und Probenqualität, Standardisierung und die Einhaltung verschiedener Regularien stetig ansteigen, nicht zuletzt deshalb, weil in immer größeren Konsortien geforscht wird. Auf die Notwendigkeit der Investition in medizinische Infrastrukturen und die wachsende Relevanz von Wissensinfrastrukturen, wie sie die TMF repräsentiert, haben die Vorsitzenden von MFT und TMF in einem gemeinsamen Beitrag hingewiesen [14].

---

20 Siehe <http://risources.dfg.de/> (06.11.2016).

21 Siehe [http://www.dfg.de/european\\_charter\\_access\\_research\\_infrastructures.pdf](http://www.dfg.de/european_charter_access_research_infrastructures.pdf) (05.10.2016).

## **TMF – IT-Infrastrukturen in der patientenorientierten Forschung**

Nicht zuletzt trägt der mit dieser Publikation in die vierte Ausgabe gehende jährlich vorgelegte Bericht des IT-Reviewing-Boards der TMF zu „IT-Infrastrukturen in der patientenorientierten Forschung – Aktueller Stand und Handlungsbedarf“ zu einer kontinuierlichen Bestands- und Bedarfsanalyse der IT-Infrastrukturen in der medizinischen Forschung bei. Der Bericht gibt insbesondere großen Universitätskliniken, aber auch andern forschungsnahen Krankenhäusern ein Nachschlagewerk für ihre IT-Planung an die Hand. Er vermittelt aktuelle Best-Practice-Beispiele für den IT-Infrastrukturaufbau in verschiedenen Forschungsprojekten und gibt Hinweise auf Tools, die sich in der Anwendung bereits bewährt haben. Dabei werden insbesondere auch Lücken identifiziert, und es wird aufgezeigt, wo dringender Handlungsbedarf für die Förderung von Forschungsinfrastrukturen besteht. Der erste Bericht [15] hatte vor allem Primärdatenquellen untersucht und gezeigt, dass bereits zahlreiche IT-Komponenten für die verschiedenen Anforderungen vorliegen, die nun jedoch in die IT-Gesamtlösung von Forschungseinrichtungen und -verbänden integriert werden müssen. Der zweite Bericht [16] lenkt den Blick zusätzlich auf Sekundärdatenquellen sowie auf die so genannten „Omics“-Daten. In der dritten Ausgabe wurden in zusätzlichen Kapiteln die Themenbereiche Anonymisierung von Patienten- oder Probandendaten, Cloud-Computing, Forschungsdatenmanagement sowie Strategische Aktivitäten und Fördermaßnahmen im Bereich der IT-Infrastrukturen analysiert [17]. Im hier vorgelegten vierten Bericht sind die Kapitel jeweils um neue Entwicklungen erweitert und aktualisiert worden.

## **ToolPool Gesundheitsforschung**

Die TMF hat in den letzten Jahren im Rahmen des vom BMBF geförderten Projekts MethInfraNet (Maßnahmen zur methodischen und infrastrukturellen Vernetzung für Qualitäts- und Effizienzsteigerung in der medizinischen Forschung) das Web-Portal „ToolPool Gesundheitsforschung“ zur Unterstützung des Aufbaus und Betriebs von IT-Infrastrukturen in der klinischen Forschung konzipiert und entwickelt. Das Ziel ist, die IT-Infrastrukturen, die für medizinische Forschung benötigt werden, und damit in Zusammenhang stehende Themen stärker ins Bewusstsein der Akteure (Forscher, Medizininformatiker, Wissenschaftsmanager, Förderorganisationen) zu rücken und konkrete Anleitung und Unterstützung anzubieten. Das Portal bündelt Produkte (Software, eServices, Berichte, Konzepte, Arbeitsmaterialien, Checklisten etc.) verschiedenster Anbieter und bietet leicht verständliche Einführungen in die beim Aufbau und Betrieb von IT-Infrastrukturen in der medizinischen Forschung entstehenden Anforderungen. Die Auswahl präsentierter Lösungen erfolgt nach transparenten und in der Community der TMF entwickelten Kriterien. Die thematische Struktur des Portals ist an diejenige

der jährlichen Berichte des IT-Reviewing-Boards in der TMF angelehnt, so dass eine enge inhaltliche Verzahnung dieser beiden Angebote erreicht werden kann. Der Start des Betriebs ist für Anfang 2017 geplant.<sup>22</sup>

### **Forschungsdatenmanagement – Leitlinien zum Umgang mit Forschungsdaten**

Während die Autoren im Kapitel 10 „Forschungsdatenmanagement“ ausführlich auf die IT-Prozesse und -Werkzeuge eingehen, die aktuell im Forschungsdatenmanagement als State-of-the-art angesehen werden können und an einigen Standorten in Deutschland bereits eingesetzt werden, gibt der folgende Abschnitt eher einen generellen Überblick zum Thema Forschungsdatenmanagement und geht dabei auch auf allgemeine Leitlinien zum Umgang mit Forschungsdaten ein.

Forschungsdaten bilden einen Grundpfeiler des wissenschaftlichen Erkenntnisgewinns. Das Forschungsdatenmanagement, d. h. die Dokumentation nicht nur der Forschungsdaten selbst, sondern auch aller für die Erhebung, Speicherung, Auswertung und Archivierung relevanten Prozesse nach fachspezifischen Standards ist daher nicht nur bedeutsam für die Qualitätssicherung wissenschaftlicher Arbeit, sondern insbesondere auch eine grundlegende Voraussetzung für die Nachvollziehbarkeit und Nachnutzbarkeit von Forschungsergebnissen. Am 30. September 2015 verabschiedete der Senat der Deutschen Forschungsgemeinschaft „Leitlinien zum Umgang mit Forschungsdaten“, in denen die Erwartungen der DFG an Antragstellende hinsichtlich des Umgangs mit den in den Forschungsprojekten entstehenden Daten formuliert werden [18]. Der Senat der DFG betont, dass die Leitlinien lediglich einen übergeordneten Handlungsrahmen definieren, dessen konkrete fachspezifische Ausgestaltung in Eigenverantwortung der unterschiedlichen Wissenschaftsbereiche geleistet werden müsse. Mit der Veröffentlichung der Leitlinien ist daher ein Appell an die Fachgemeinschaften verbunden, ihren Umgang mit Forschungsdaten zu reflektieren und zu definieren, welche der im jeweiligen Feld entstehenden Forschungsdaten besonders wertvoll sind und in welcher Weise die Möglichkeiten für deren Nachnutzung – zum Beispiel durch eine qualitätsgesicherte Speicherung in einem geeigneten Forschungsdatenrepositorium – verbessert werden können.

Weitere Empfehlungen zum Forschungsdatenmanagement hat die Mitgliederversammlung der Hochschulrektorenkonferenz (HRK) am 10.11.2015 unter dem Titel “Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können. Orientierungspfade, Handlungsoptionen, Szenarien” verabschiedet [19]. Diese bauen

---

22 Siehe [www.toolpool-gesundheitsforschung.de](http://www.toolpool-gesundheitsforschung.de).

auf einer vorhergehenden Empfehlung der HRK aus dem Jahr 2014 auf [20]. Damals wurde die strategische Bedeutung des Themas für die Hochschulen in Deutschland betont. Die jetzt veröffentlichten Empfehlungen richten sich darüber hinaus auch an die Politik, indem die Verantwortlichen in Bund und Ländern aufgefordert werden, sowohl eine koordinierende Rolle beim Aufbau übergreifender Infrastrukturen für das Forschungsdatenmanagement zu übernehmen als auch die notwendige Finanzierung zu sichern. Die HRK sieht aufgrund der Vielzahl der Akteure einen dringenden Bedarf an Koordination und Abstimmung – auch zwischen den Hochschulen – beim Aufbau eines länderübergreifenden und international kompatiblen Forschungsdatenmanagements. Dabei müsse das Ziel sein, möglichst alle Hochschulen einzubinden.

Auch die Allianz der deutschen Wissenschaftsorganisationen hat 2015 ein Positionspapier zum Umgang mit Forschungsdaten publiziert [21]. Die Arbeitsgruppe Forschungsdaten der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen<sup>23</sup> stellt in diesem Arbeitspapier die aktuelle Situation im Themenfeld dar und benennt offene Herausforderungen, um damit eine Grundlage für notwendige Diskussionen und Überlegungen zum zukünftigen Umgang mit Forschungsdaten zu schaffen. Grundlage dafür sind die „Grundsätze zum Umgang mit Forschungsdaten“, die von der Allianz der deutschen Wissenschaftsorganisationen 2010 verabschiedet wurden.<sup>24</sup>

Ein sehr umfassendes Empfehlungspapier zum Forschungsdatenmanagement hat darüber hinaus im Mai 2016 der Rat für Informationsinfrastrukturen vorgelegt. Dessen Inhalte wurden im Abschnitt „Rat für Informationsinfrastrukturen“ weiter oben in diesem Kapitel bereits ausführlich dargestellt.

## Fördermaßnahmen im Bereich der IT-Infrastrukturen in Deutschland

### **Strukturmaßnahmen des BMBF in der medizinischen Forschung**

Im Bereich der sogenannten klinisch-evaluativen Forschung förderte und fördert das BMBF seit 1999 sowohl die Kompetenznetze in der Medizin als auch die Koordinierungszentren für klinische Studien bzw. seit 2007 die Klinischen Studienzentren. Parallel dazu wurde die TMF zunächst als Projekt und später über Mitgliedsbeiträge sowie das Vorhaben MethInfraNet gefördert. Durch diese Maßnahmen hat das BMBF erfolgreiche Strukturen und Netzwerke in der medizinischen Forschung etabliert. Weiter unterstützt wurden diese Strukturmaßnahmen durch den Förderschwerpunkt

---

23 Siehe <http://www.allianzinitiative.de/> (06.11.2016).

24 Siehe <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze.html> (06.11.2016).

„Förderung von Instrumenten- und Methodenentwicklung für die patientenorientierte medizinische Forschung“, der von 2008 – 2014 durchgeführt wurde und mit dem Einzelprojekte zu methodischen, informationstechnologischen, organisatorischen und qualitätssichernden Fragestellungen gefördert wurden. Seit 2009 fördert das BMBF mit der Gründung der sechs Deutschen Zentren der Gesundheitsforschung (DZG) auch Zentren und Institute. Alle diese Fördermaßnahmen umfassen in verschiedenem Umfang auch IT-Infrastrukturmaßnahmen. Insbesondere in Bezug auf die DZG kann davon ausgegangen werden, dass die sich dort im Aufbau befindlichen IT-Architekturen in gewissem Maße Vorarbeiten für die weiter unten diskutierte Medizininformatik-Initiative sein werden. Im Oktober 2016 hat das BMBF eine Bekanntmachung zur Förderung des Aufbaus modellhafter patientenbezogener Register für die Versorgungsforschung publiziert. Um eine hohe Qualität der aufzubauenden Register zu erreichen sowie um registerübergreifend methodische, technische und strukturelle Standards gewährleisten, soll parallel ein Begleitprojekt gefördert werden.

### **BMBF-Förderkonzept Bioinformatik**

Im März 2015 wurde das „Deutsche Netzwerk für Bioinformatik-Infrastruktur“ (de. NBI)<sup>25</sup> als neues Großprojekt im Umfang von mehr als 22 Millionen Euro Fördersumme ins Leben gerufen. Das Deutsche Netzwerk für Bioinformatik-Infrastruktur besteht aus einer Koordinierungseinheit und acht Leistungszentren, die über spezifische Expertisen und Ressourcen in der Bioinformatik verfügen. Unterstützt durch die Koordinierungseinheit wird ein Netzwerk aufgebaut, das die Kompetenzen der einzelnen Leistungszentren bündelt, Datenressourcen und bioinformatische Werkzeuge besser verfügbar machen und Dienstleistungen anbieten soll. Weiterhin werden Schulungen, Workshops und Summer Schools durchgeführt. Das Projekt geht auf eine Empfehlung des BioÖkonomieRats aus dem Jahr 2012 zurück<sup>26</sup> und fungiert seit 2016 mit der Ratifizierung des ELIXIR Konsortialvertrags durch das BMBF auch als deutscher Knoten zur europäischen ESFRI-Infrastruktur ELIXIR.

### **BMBF-Förderkonzept Medizininformatik**

Am 16.11.2015 wurde die „Richtlinie zur Förderung der Konzeptphase sowie der Aufbau- und Vernetzungsphase im Förderkonzept Medizininformatik“ bekannt gegeben. Diese Bekanntmachung baut auf dem Förderkonzept Medizininformatik [22] auf, mit der auf förderpolitischer Ebene die Medizin und die Informationstechnologie zusammengeführt, die bio- und medizininformatische Forschung gestärkt und eine datenbasierte

---

25 Siehe <https://www.denbi.de/>, (06.11.2016).

26 Siehe [https://www.ptj.de/1w\\_resource/datapool/\\_items/item\\_4718/boer\\_broschuere\\_bioinformatik.pdf](https://www.ptj.de/1w_resource/datapool/_items/item_4718/boer_broschuere_bioinformatik.pdf) (06.11.2016).

Patientenversorgung etabliert werden sollen. Grundlegendes Ziel des Förderkonzepts ist es, den Austausch und die Nutzung von Daten aus Krankenversorgung, klinischer und biomedizinischer Forschung über die Grenzen von Institutionen und Standorten hinweg zu ermöglichen. Im Fokus stehen dabei zunächst die Universitätskliniken, wobei die Anwendung über die Universitätskliniken hinaus jedoch von Beginn an mitbedacht werden soll. Das Förderkonzept ist modular aufgebaut und verläuft über mehrere Ausbaustufen. Wenn die ersten Stufen erfolgreich verlaufen, plant das BMBF, sich längerfristig und mit substanziellen Fördermitteln zu engagieren.

Seit August 2016 fördert das BMBF in der Konzeptphase sieben Konsortien. Sie sollen Strategien für die gemeinsame Datennutzung und den Datenaustausch im Konsortium und darüber hinaus entwickeln und konkrete medizinische Anwendungsfälle ausarbeiten. Die Konsortien umfassen jeweils mindestens zwei Universitätskliniken sowie weitere Krankenhäuser, medizinische Forschungseinrichtungen und Industrieunternehmen. Ein Nationales Steuerungsgremium koordiniert die Zusammenarbeit der Konsortien. Um die Zusammenarbeit und Abstimmung der Konsortien im Nationalen Steuerungsgremium organisatorisch und fachlich zu unterstützen, wird ein Begleitprojekt gefördert. Dieses wird von der TMF zusammen mit dem Medizinischen Fakultätentag (MFT) und dem Verband der Universitätsklinika Deutschlands (VUD) durchgeführt.

## **DFG – Großgeräteprogramme**

Bereits seit Jahrzehnten existieren Investitionsfördermöglichkeiten für Großgeräte an Hochschulen. Die DFG war zunächst im Rahmen des Hochschulbaufördergesetzes (HBFVG) und ist seit der Föderalismusreform 2007 in Folgeprogrammen in die Begutachtung und Bewertung von Großgeräteanträgen eingebunden, die gemeinschaftlich von Bund und Ländern finanziert werden. Einen Überblick über die Programme, ihre Charakteristika und Nutzung ist in einem 5-Jahresbericht (2007-2011)<sup>27</sup> zu finden. Jährlich werden über diese Programme ca. 400 Mio. Euro, verteilt auf einige hundert Projekte, in die Geräteinfrastruktur deutscher Hochschulen investiert. Mit Fokus auf die Universitätsmedizin ist insbesondere das Programm „Großgeräte der Länder“ von Interesse, in welchem neben Hardware, Speichersystemen etc. auch IT-Systeme (KIS, Subsysteme etc.) als Antragsgegenstand möglich sind. Die DFG sieht sich hier zunehmend auch in einer Beratungsfunktion, sind die Anträge doch häufig integrale Bestandteile eines IT-Konzepts mit Auswirkungen auf die strategische Ausrichtung eines Universitätsklinikums. Darüber hinaus eröffnet das Programm „Forschungsgroßgeräte“ Finanzierungsmöglichkeiten für forschungsbezogene Geräte oder IT-Systeme. Die weiter oben angeführten Empfehlungen der Kommission für IT-Infrastruktur geben unmittelbare

---

<sup>27</sup> Siehe [http://www.dfg.de/foerderung/programme/wgi/fuenf\\_jahre\\_grossgeraeteprogramme.pdf](http://www.dfg.de/foerderung/programme/wgi/fuenf_jahre_grossgeraeteprogramme.pdf) (06.11.2016).

Unterstützung für die Wahl des geeigneten Förderprogramms sowie für die Schaffung adäquater Rahmenbedingungen, die für einen erfolgreichen Betrieb essentiell sind.

Die Großgeräteprogramme sind offen, Anträge können jederzeit eingereicht werden. Sie liefern damit einen essentiellen Beitrag für eine hochwertige Forschungsinfrastruktur an deutschen Hochschulen und Universitätsklinika.

### **DFG – Informationsinfrastrukturen für Forschungsdaten**

2013 startete der Hauptausschuss der DFG ein neues Förderprogramm mit dem Ziel, qualitätsgesicherte Forschungsdaten, die häufig mit großem finanziellen und zeitlichen Aufwand erhoben werden, nachhaltiger zu sichern und damit für die wissenschaftliche Nachnutzung verfügbar zu machen. Das Förderprogramm ist eine konkrete Antwort der DFG auf die oben genannten Diskussionen und Empfehlungen zum digitalen Wandel und dessen Auswirkungen auf Informationsinfrastrukturen. Das neue Förderangebot der DFG orientiert sich somit an dem Bedarf der Wissenschaft, Strukturen für Forschungsdaten zu entwickeln und zu verstetigen. Gefördert werden können Projekte zur Entwicklung tragfähiger Konzepte und Lösungen für überregionale Dateninfrastrukturen für eine oder mehrere Fachdisziplinen bis hin zum weiteren Ausbau bereits existierender Infrastrukturen, beispielsweise mit Blick auf internationale Standards und eine grenzüberschreitende Zusammenarbeit.

### **DFG – Förderung von Forschungsprojekten über und mit der Technologie- und Methodenplattform für die vernetzte Medizinische Forschung (TMF)**

In einigen der oben genannten Empfehlungen von Wissenschaftsrat [2] und DFG [12] wird die TMF als eine der Begleitstrukturen für die medizinische Forschung benannt, die in Bezug auf die notwendigen IT-Infrastrukturen, die Rahmenbedingungen in den Bereichen Ethik und Datenschutz sowie bei der Erarbeitung strategischer Empfehlungen und Richtlinien wertvolle Vorarbeiten geleistet hat. Mit ihrer Ausschreibung zur „Förderung von Forschungsprojekten über und mit der Technologie- und Methodenplattform für vernetzte Medizinische Forschung (TMF)“ von 2015 greift die DFG diese Sichtweise auf und führt sie fort. Ziel der Fördermaßnahme ist es, die Nützlichkeit der aktuellen Instrumente und Angebote der TMF für die medizinische Forschung zu erproben sowie neue Ansätze für die zukünftige Arbeit der TMF zu entwickeln. In einem offenen Workshop konnten Antragsteller ihre Projektideen miteinander diskutieren und in einer zweiten Stufe ausgearbeitete Projektanträge bei der DFG einreichen. Seit April 2016 fördert die DFG acht Projekte, die Infrastrukturen für die medizinische Forschung in Zusammenarbeit mit der TMF aufbauen, sowie ein TMF-Zentralprojekt.

## Innovationsfonds

Nachdem sich im Oktober 2015 der Innovationsausschuss beim Gemeinsamen Bundesausschuss (G-BA) konstituiert hat, ist 2016 auch der von der Bundesregierung neu aufgelegte Innovationsfonds gestartet.<sup>28</sup> Dessen Fördersumme beträgt in den Jahren 2016 bis 2019 jeweils 300 Millionen Euro jährlich. Hierbei sind 225 Millionen Euro für die Förderung neuer Versorgungsformen und 75 Millionen Euro für die Versorgungsforschung vorgesehen. Die ersten Bewilligungen sind im Oktober 2016 erfolgt. Zum jetzigen Zeitpunkt kann noch nicht abgesehen werden, inwieweit die dort geförderten Projekte Auswirkungen auf die IT-Infrastrukturen in der medizinischen, patientenorientierten Forschung haben werden. Sicher ist jedoch, dass die Sekundärnutzung von Versorgungsdaten eines der zentralen Themen sein wird und hierzu viele neue Anwendungsfälle zu erwarten sind. Somit ist zu hoffen, dass vor dem Hintergrund dieser Fördermaßnahme weitere Schritte hin auf einen einheitlichen und adäquaten Rechtsrahmen für die Nutzung von Versorgungsdaten der behandelnden Einrichtungen einerseits und der Leistungsträger andererseits möglich sein werden.

## Bewertung und Handlungsbedarf

Wie sich aus den obigen Ausführungen ableiten lässt, wird die Notwendigkeit der Bereitstellung von leistungsfähigen IT-Infrastrukturen für eine erfolgreiche medizinische Forschung von der Wissenschaftspolitik, ihren Expertengremien und den Förderern inzwischen gleichermaßen anerkannt. DFG und BMBF haben Förderprogramme und -konzepte entwickelt, die der Notwendigkeit von langfristigen und nachhaltigen Lösungen für die Wissenschaft allgemein bzw. auch spezifisch für die Medizininformatik Rechnung tragen. Man darf hoffen, dass hiermit ein erster Schritt getan ist, um von der projektgetriebenen, kurzfristigen Förderung von Forschungsinfrastrukturen wegzukommen, die in den letzten Jahren zu einer Vielzahl von – in der Regel nicht kompatiblen – Insellösungen in der Universitätsmedizin geführt hat. Es wird weiterhin dringender Handlungsbedarf gesehen, die Entwicklung von strategisch durchdachten, standardisierten und miteinander kompatiblen IT-Infrastrukturen konsequent weiter zu führen, wie dies auch im Empfehlungspapier des RfII formuliert ist [6].

Eine weitere große Herausforderung, die in den nächsten Jahren adressiert werden muss, ist die Finanzierung der Betriebskosten von Forschungsinfrastrukturen angesichts knapper werdender Budgets der öffentlichen Haushalte. Auch wenn die zentrale Finanzierung der Investitionskosten von Forschungsinfrastrukturen als notwendig anerkannt

---

28 Siehe <https://innovationsfonds.g-ba.de/> (06.11.2016).

ist, Belege hierfür sind der Roadmap-Prozess sowie die genannten Förderprogramme, ist das Problem der Betriebskosten damit noch nicht gelöst. Auch die inzwischen 10-jährige Archivierungspflicht von Forschungsprimärdaten bei stark wachsenden Datenvolumina gehört in diese Betrachtungen hinein. Obwohl als Kostenfaktor erkannt, können beispielsweise innerhalb der aktuellen BMBF-Ausschreibung für Forschungsinfrastrukturen zur Aufnahme in die Nationale Roadmap Betriebskosten, die während der Nutzungsphase der Forschungsinfrastruktur anfallen und nicht entscheidend zur Weiterentwicklung von Daten- oder Informationssammlungen beitragen, nicht angesetzt werden. Auch das Förderkonzept Medizininformatik des BMBF adressiert die Betriebskosten nur insofern, als dass entsprechende Zusagen seitens der Universitätskliniken und ihrer Träger für die langfristige Fortführung der aufzubauenden Strukturen verlangt werden. Hier ist absehbar, dass man zu gemeinsamen, übergreifenden Konzepten von Bund und Ländern und den von ihnen getragenen universitären und außeruniversitären Forschungseinrichtungen bzw. -organisationen kommen muss. Im Rahmen der Allianz der Wissenschaftsorganisationen wird dieser Themenkomplex eine gewichtige Rolle spielen.

Zum Betrieb von Forschungsinfrastrukturen gehört nicht zuletzt auch die Bereitstellung der dafür notwendigen personellen Ressourcen. Schon durch einen Blick auf die Kapitel in diesem Report lässt sich belegen, dass hierfür Personal mit hoher methodischer Kompetenz notwendig ist. Dem steht in weiten Bereichen der Medizininformatik, Biometrie, Epidemiologie und Bioinformatik in Deutschland ein Mangel an geeigneten Nachwuchskräften gegenüber. Hier muss mit geeigneten Aus- und Weiterbildungsangeboten gegengesteuert werden.

## Literatur

1. WR *Übergreifende Empfehlungen zu Informationsinfrastrukturen*. 2011. Wissenschaftsrat, Drs. 10466-11, <http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf> (Abruf: 2016-12-01).
2. WR *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020*. 2012. Wissenschaftsrat, Drs. 2359-12, <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf> (Abruf: 2016-12-01).
3. GWK *Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder*. 2011. Kommission Zukunft der Informationsinfrastruktur, [http://www.leibniz-gemeinschaft.de/fileadmin/user\\_upload/downloads/Infrastruktur/KII\\_Gesamtkonzept.pdf](http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf) (Abruf: 2016-12-01).
4. BLK *Neuausrichtung der öffentlich geförderten Informationseinrichtungen. Abschlussbericht der BLK-Arbeitsgruppe „Zukunft der Fachinformation“*. 2006. Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung (BLK), Heft 138, ISBN Nr. 3-934850-85-5, <http://www.blk-bonn.de/papers/heft138.pdf> (Abruf: 2016-12-01).
5. WR *Perspektiven der Universitätsmedizin*. 2016. Wissenschaftsrat, Drs. 5663-16, <http://www.wissenschaftsrat.de/download/archiv/5663-16.pdf> (Abruf: 2016-10-26).
6. RfII *Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland*. 2016. Rat für Informationsinfrastrukturen, <http://www.rfii.de/?wpdmdl=1998> (Abruf: 2016-10-10).
7. EC *Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud*. 2016. European Commission, [http://ec.europa.eu/research/openscience/pdf/realising\\_the\\_european\\_open\\_science\\_cloud\\_2016.pdf](http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf) (Abruf: 2016-12-01).
8. DFG *Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Stellungnahme der Kommission für IT-Infrastruktur für 2016–2020*. 2016. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/foerderung/programme/wgi/kfr\\_stellungnahme\\_2016\\_2020.pdf](http://www.dfg.de/download/pdf/foerderung/programme/wgi/kfr_stellungnahme_2016_2020.pdf) (Abruf: 2016-11-23).
9. DFG *Cloud-Dienste - Addendum zu den Empfehlungen der Kommission für IT-Infrastruktur 2011-2015*. 2014. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/foerderung/programme/wgi/addendum\\_cloud\\_dienste\\_kfr\\_2014.pdf](http://www.dfg.de/download/pdf/foerderung/programme/wgi/addendum_cloud_dienste_kfr_2014.pdf) (Abruf: 2016-02-09).

10. DFG *Positionspapier der Deutschen Forschungsgemeinschaft: Die digitale Transformation weiter gestalten – Der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung*. 2012. Deutsche Forschungsgemeinschaft, Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme [http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier\\_digitale\\_transformation.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_digitale_transformation.pdf) (Abruf: 2016-12-01).
11. DFG *Strukturelle Rahmenbedingungen der klinischen Forschung in der deutschen Universitätsmedizin. Stellungnahme der Ständigen Senatskommission für Grundsatzfragen in der klinischen Forschung der Deutschen Forschungsgemeinschaft*. 2014. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/2014/140610\\_stellungnahme\\_klin\\_forschung.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2014/140610_stellungnahme_klin_forschung.pdf) (Abruf: 2016-02-09).
12. DFG *Zur Weiterentwicklung der Klinischen Forschung an der deutschen Universitätsmedizin in den Jahren 2015 – 2025. Empfehlungen der Ständigen Senatskommission für Grundsatzfragen in der Klinischen Forschung der Deutschen Forschungsgemeinschaft*. 2015. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/2015/sgkf\\_empfehlungen\\_klinische\\_forschung\\_150720.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/2015/sgkf_empfehlungen_klinische_forschung_150720.pdf) (Abruf: 2016-02-09).
13. Kroemer, H.K., Hrsg. *Infrastrukturen und Rahmenbedingungen für die medizinische Forschung. Gemeinsames Symposium des MFT Medizinischen Fakultätentages der Bundesrepublik Deutschland und der TMF - Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.* 1. Aufl. 2012, MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland e.V., Berlin.
14. Kroemer, H.K., Krawczak, M., *Forschung braucht Infrastrukturen*. Deutsches Ärzteblatt, 2014. **111**(5): S. A170 - A171.
15. Drepper, J., Semler, S.C., Hrsg. *IT-Infrastrukturen in der patientenorientierten Forschung. Aktueller Stand und Handlungsbedarf – 2012/2013. Verfasst und vorgelegt vom IT-Reviewing-Board der TMF*. 2014, AKA, Berlin.
16. Drepper, J., Semler, S.C., Hrsg. *IT-Infrastrukturen in der patientenorientierten Forschung. Aktueller Stand und Handlungsbedarf – 2014. Verfasst und vorgelegt vom IT-Reviewing-Board der TMF*. 2015, AKA, Berlin.
17. Drepper, J., Semler, S.C., Hrsg. *IT-Infrastrukturen in der patientenorientierten Forschung. Aktueller Stand und Handlungsbedarf – 2015. Verfasst und vorgelegt vom IT-Reviewing-Board der TMF*. 2016, AKA, Berlin, <http://www.tmf-ev.de/Produkte/ITReport>.

18. DFG *Leitlinien zum Umgang mit Forschungsdaten*. 2015. Deutsche Forschungsgemeinschaft, [http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien\\_forschungsdaten.pdf](http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf) (Abruf: 2016-12-01).
19. HRK *Empfehlung der 19. Mitgliederversammlung der HRK am 10. November 2015 in Kiel: Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können. Orientierungspfade, Handlungsoptionen, Szenarien*. 2015. Hochschulrektorenkonferenz, [https://www.hrk.de/uploads/tx\\_szconvention/Empfehlung\\_Forschungsdatenmanagement\\_\\_final\\_Stand\\_11.11.2015.pdf](https://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement__final_Stand_11.11.2015.pdf) (Abruf: 2016-12-01).
20. HRK *Empfehlung der 16. Mitgliederversammlung der HRK am 13. Mai 2014 in Frankfurt am Main: Management von Forschungsdaten – eine zentrale strategische Herausforderung für Hochschulleitungen*. 2014. Hochschulrektorenkonferenz, [https://www.hrk.de/uploads/tx\\_szconvention/HRK\\_Empfehlung\\_Forschungsdaten\\_13052014\\_01.pdf](https://www.hrk.de/uploads/tx_szconvention/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf) (Abruf: 2016-12-01).
21. Franke, M., Heinzl, S., Mauer, R. et al. *Positionspapier „Research data at your fingertips“ der Arbeitsgruppe Forschungsdaten im Rahmen der Schwerpunktinitiative „Digitale Information“ der Allianz der Wissenschaftsorganisationen*. 2015. GeoForschungsZentrum GFZ, <http://doi.org/10.2312/allianzfd.001> (Abruf: 2016-02-09).
22. BMBF *Förderkonzept Medizininformatik. Daten vernetzen – Gesundheitsversorgung verbessern*. 2015. Bundesministerium für Bildung und Forschung (BMBF), Referat Methoden- und Strukturentwicklung in den Lebenswissenschaften [http://www.gesundheitsforschung-bmbf.de/\\_media/BMBF\\_040\\_Medizininformatik\\_BARRIEREFREI.pdf](http://www.gesundheitsforschung-bmbf.de/_media/BMBF_040_Medizininformatik_BARRIEREFREI.pdf) (Abruf: 2016-12-01).



# Abkürzungsverzeichnis

AAL	Ambient Assisted Living
ACGT	Advancing Clinico Genomic Trials on Cancer ( <a href="http://acgt.ercim.eu">http://acgt.ercim.eu</a> )
ACM	Association for Computing Machinery ( <a href="http://www.acm.org">www.acm.org</a> )
ACTION	Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks; PPP mit der FDA zur Verbesserung der Forschung und Unterstützung klinischer Studien im angegebenen Krankheitsbereich ( <a href="http://www.action.org">www.action.org</a> )
ADaM	Analysis Dataset Model (CDISC-Standard)
ADOPT	implemEntation and OPeration of the gateway for healTh into BBMRI-ERIC
ADR	Alternative Dispute Resolution – Außergerichtliche Streitbeilegung
ADT	Arbeitsgemeinschaft Deutscher Tumorzentren e.V. ( <a href="http://www.tumorzentren.de">www.tumorzentren.de</a> )
AG	Aktiengesellschaft
AG	Arbeitsgruppe
AGPL	Affero General Public License
AGPLv3	GNU Affero General Public License Version 3
AHRQ	Agency for Healthcare Research and Quality des US Department of Health & Human Services ( <a href="http://www.ahrq.gov">www.ahrq.gov</a> )
AIDS	Acquired Immune Deficiency Syndrome: Krankheitsbild einer erworbenen Immunschwäche aufgrund einer HIV-Infektion
AIS	Arztinformationssystem
AKmedEK	Arbeitskreis Medizinischer Ethik-Kommissionen in der Bundesrepublik Deutschland ( <a href="http://www.ak-med-ethik-komm.de">www.ak-med-ethik-komm.de</a> )
AMG	Gesetz über den Verkehr mit Arzneimitteln – Arzneimittelgesetz
AMIA	American Medical Informatics Association ( <a href="http://www.amia.org">www.amia.org</a> )
AnonTool	Anonymisierungstool der TMF ( <a href="http://www.tmf-ev.de/Produkte/P100201">www.tmf-ev.de/Produkte/P100201</a> )
ANZICS	Australian and New Zealand Intensive Care Society ( <a href="http://www.anzics.com.au">www.anzics.com.au</a> )
AP	Arbeitspaket
APACHE	Acute Physiology And Chronic Health Evaluation
API	Application Programming Interface: Softwareschnittstelle zu Programmen oder Betriebssystemen
APIS	Arztpraxisinformationssystem
APRE	Agency for the Promotion of European Research ( <a href="http://www.apre.it/en">www.apre.it/en</a> )
AQUA	Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen GmbH ( <a href="http://www.aqua-institut.de">www.aqua-institut.de</a> )
ARCH	Archivierung
ARX	Data Anonymization Tool der Technischen Universität München ( <a href="http://arx.deidentifier.org">http://arx.deidentifier.org</a> )

ASCII	American Standard Code for Information Interchange; standardisierter Zeichensatz, der von den meisten Computersystemen interpretiert werden kann.
ASP	Application Service Provider
ATGC	Applied & Translational Genomics Cloud ( <a href="http://www.applied-translational-genomics-cloud.de">www.applied-translational-genomics-cloud.de</a> )
AUS	Datenanalyse oder -verwertung
BAM	Compressed Binary SAM-Format
BAnz AT	Bundesanzeiger, Amtlicher Teil
BBMRI	[European] Biobanking and Biomolecular Resources Research Infrastructure, ein ESFRI-Projekt ( <a href="http://www.biobanks.eu">www.biobanks.eu</a> )
BBMRI-ERIC	Europäische Biobanken-Infrastruktur ( <a href="http://bbmri.eu">http://bbmri.eu</a> )
BBMRI-LPC	Biobanking and Biomolecular Resources Research Infrastructure – Large Prospective Cohorts; Netzwerk großer prospektiver Kohorten in BBMRI
BCL	Base Call; binäres Dateiformat für base calling und Qualitätsinformationen zu genetischen Sequenzierungsdaten
BDSG	Bundesdatenschutzgesetz
BGB	Bürgerliches Gesetzbuch
BIH	Berliner Institut für Gesundheitsforschung / Berlin Institute of Health; Gemeinsame Forschungseinrichtung der Charité und des MDC ( <a href="http://www.bihealth.org">www.bihealth.org</a> )
Bioconductor	Open Source Software Projekt zur Analyse genetischer Daten auf Basis der Programmiersprache R ( <a href="http://www.bioconductor.org">www.bioconductor.org</a> )
BioDAS	Distributed Annotation System; Kommunikationsprotokoll für Annotationsdaten zu Gen- oder Proteinsequenzen ( <a href="http://www.biodas.org">www.biodas.org</a> )
BioGRID	Biological General Repository for Interaction Datasets ( <a href="http://thebiogrid.org">http://thebiogrid.org</a> )
BioMart	Datenbanksystem für biologische Daten ( <a href="http://www.biomart.org">www.biomart.org</a> )
BioMedBridges	EU-Projekt zur Entwicklung gemeinsamer, harmonisierter Lösungen für die biomedizinischen ESFRI-Infrastrukturen
BioModels	Datenbank für Modelle von biologischen Prozessen ( <a href="http://www.ebi.ac.uk/biomodels-main">www.ebi.ac.uk/biomodels-main</a> )
BioÖkonomieRat	Bioökonomierat ( <a href="http://biooekonomierat.de">http://biooekonomierat.de</a> )
BioPortal	Datenbank zu biomedizinischen Ontologien des NCBO ( <a href="http://bioportal.bioontology.org">http://bioportal.bioontology.org</a> )
BLK	Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung ( <a href="http://www.blk-bonn.de">www.blk-bonn.de</a> )
BMBF	Bundesministerium für Bildung und Forschung ( <a href="http://www.bmbf.de">www.bmbf.de</a> )
BMC	BioMed Central; Verlag ( <a href="http://www.biomedcentral.com">www.biomedcentral.com</a> )
BMDS	Behavioral Measurement Database Services ( <a href="http://www.bmdshapi.com">www.bmdshapi.com</a> )
BMG	Bundesministerium für Gesundheit ( <a href="http://www.bmg.bund.de">www.bmg.bund.de</a> )

BMJ	Bundesministerium der Justiz und für Verbraucherschutz ( <a href="http://www.bmj.de">www.bmj.de</a> )
BMWi	Bundesministerium für Wirtschaft und Energie ( <a href="http://www.bmwi.bund.de">www.bmwi.bund.de</a> )
BQS	Bundesgeschäftsstelle Qualitätssicherung gGmbH ( <a href="http://www.bqs-online.de">www.bqs-online.de</a> )
BRIDG	Biomedical Research Integrated Domain Group ( <a href="http://www.bridgmodel.org">www.bridgmodel.org</a> )
BRISQ	Biospecimen Reporting for Improved Study Quality
BSI	Bundesamt für Sicherheit in der Informationstechnik ( <a href="http://www.bsi.de">www.bsi.de</a> )
BVMI	Berufsverband Medizinischer Informatiker e.V. ( <a href="http://www.bvmi.de">www.bvmi.de</a> )
caBIG	cancer Biomedical Informatics Grid: Grid-Initiative des NCI ( <a href="http://cabig.cancer.gov">http://cabig.cancer.gov</a> )
caCORE	cancer Common Ontologic Representation Environment
caDSR	Cancer Data Standards Repository: Datenbank und Toolset des NCI zur Entwicklung und Verwaltung standardisierter Metadaten ( <a href="http://ncicb.nci.nih.gov/ncicb/infrastructure/cacore_overview/cadsr">http://ncicb.nci.nih.gov/ncicb/infrastructure/cacore_overview/cadsr</a> )
CancerGrid	Vom MRC gefördertes Projekt zur Verbesserung der klinischen Krebsforschung ( <a href="https://cancergrid.org">https://cancergrid.org</a> )
CARELIS	CARLOS Record Linkage System; Software des Offis Oldenburg zum Abgleich von Meldungen an Krebsregister
CARESS	CARLOS Epidemiological and Statistical Data Exploration System; epidemiologische Auswertungskomponente des Offis Oldenburg
CARLOS	Cancer Registry Lower-Saxony; Projekt zum Aufbau des Epidemiologischen Krebsregisters Niedersachsen
CARTools	Software-Suite des OFFIS Oldenburg für das Niedersächsische Krebsregister, bestehend aus den Anwendungen CARELIS und CARESS
caTISSUE	caBIG tissue bank repository tool for biospecimen inventory, tracking, and basic annotation ( <a href="https://wiki.nci.nih.gov/display/caTissue">https://wiki.nci.nih.gov/display/caTissue</a> )
CAU	Christian-Albrechts-Universität zu Kiel ( <a href="http://www.uni-kiel.de">www.uni-kiel.de</a> )
cBioPortal	Offenes Portal zur interaktiven Exploration multidimensionaler genetischer Datensätze aus der Onkologie ( <a href="http://www.cbioportal.org">www.cbioportal.org</a> )
cBMB	zentralisierte Biobank, gefördert vom BMBF im Rahmen der Nationalen Biobanken-Initiative
C-Path	Critical Path Institute ( <a href="https://c-path.org">https://c-path.org</a> )
CD	Compact Disc
CDA	Clinical Document Architecture, HL7-Standard für den Austausch klinischer Dokumente
CDASH	Clinical Data Acquisition Standards Harmonization, CDISC-Initiative
CDE	Common Data Elements
CDISC	Clinical Data Interchange Standards Consortium ( <a href="http://www.cdisc.org">www.cdisc.org</a> )
CDM	Common Data Model(s)
CDMS	Clinical Datamanagement System

CDW	Clinical Data Warehouse
CEM	Clinical Element Models
CEN	Comité Européen de Normalisation, Europäisches Komitee für Normung ( <a href="http://www.cenorm.be">www.cenorm.be</a> )
CERN	Conseil Européen pour la Recherche Nucléaire ( <a href="http://www.cern.ch">www.cern.ch</a> )
CFAST	Initiative zur Beschleunigung der klinischen Forschung und Produktentwicklung durch Standardisierung und Tools; initiiert als Partnerschaftsprojekt von CDISC und C-Path
CFR	Code of Federal Regulations der USA ( <a href="http://www.gpoaccess.gov/cfr">www.gpoaccess.gov/cfr</a> )
CHARISMHA	Chances and Risks of Mobile Health Apps; Studie der MHH im Auftrag des BMG ( <a href="http://www.digibib.tu-bs.de/?docid=00060000">www.digibib.tu-bs.de/?docid=00060000</a> )
ChIP	Chromatin Immunoprecipitation (Sequencing)
CIHI	Canadian Institute for Health Information ( <a href="http://www.cihi.ca">www.cihi.ca</a> )
CIMI	Clinical Information Modeling Initiative ( <a href="http://www.opencimi.org">www.opencimi.org</a> )
cloud4health	im Rahmen des BMWi-Förderprogramms Trusted Cloud gefördertes Projekt zur Entwicklung und Erprobung von innovativen, sicheren und rechtskonformen Cloud-Computing-Diensten im Gesundheitsbereich ( <a href="http://www.cloud4health.de">www.cloud4health.de</a> )
COE	Council of Europe – Europarat ( <a href="http://www.coe.int">www.coe.int</a> )
COPD	Chronic Obstructive Pulmonary Disease - Chronisch obstruktive Lungenerkrankung
CORBEL	Coordinated Research Infrastructures Building Enduring Life-science Services ( <a href="http://www.corbel-project.eu">www.corbel-project.eu</a> )
CRF	Case Report Form
CRM	Customer Relationship Management
CRO	Contract Research Organisation
CT	Computer-Tomografie
CTC	Clinical Trial Center (Aachen)
CTK	Common Toolkit; Open-Source-Software für die Verarbeitung von Bilddaten auf der Basis des DICOM-Standards ( <a href="http://www.commonstk.org">www.commonstk.org</a> )
CTMS	Clinical Trial Management System
CTP	RSNA Clinical Trials Processor
CTS	Common Terminology Services, HL7-Spezifikation zum Funktionsumfang eines Terminologieservers
CTS2	Common Terminology Services 2.0, Erweiterung der ursprünglichen HL7-Spezifikation CTS
D-A-CH	Kürzel für die Länder Deutschland (D), Österreich (A) und Schweiz (CH) gemäß der Nationalitätenkennzeichen für Kraftfahrzeuge

DANA	Die Datenschutznachrichten: Zeitschrift der Deutschen Vereinigung für Datenschutz e.V. ( <a href="http://www.datenschutzverein.de">www.datenschutzverein.de</a> )
DataCite	Internationale Non-Profit-Organisation zur zitierfähigen Auszeichnung von Forschungsdaten mit Hilfe des DOI-Systems ( <a href="http://www.datacite.org">www.datacite.org</a> )
Dataverse	Open source research data repository software des IQSS ( <a href="http://dataverse.org">http://dataverse.org</a> )
DBR	Deutsches Biobanken-Register ( <a href="http://www.biobanken.de">www.biobanken.de</a> )
DC	Dublin Core (siehe DCMI)
DCM	HL7 Detailed Clinical Models
DCMI	Dublin Core Metadata Initiative ( <a href="http://dublincore.org">http://dublincore.org</a> )
DCMTK	DICOM-Toolkit des Offis Oldenburg ( <a href="http://dicom.offis.de/dcmk">http://dicom.offis.de/dcmk</a> )
DCN	Death Certificate Notification, Parameter aus der epidemiologischen Krebsregistrierung
DCO	Death Certificate Only, Parameter aus der epidemiologischen Krebsregistrierung
DDI	Data Documentation Initiative ( <a href="http://www.ddialliance.org">www.ddialliance.org</a> )
de.NBI	Deutsches Netzwerk für Bioinformatik-Infrastruktur ( <a href="http://www.denbi.de">www.denbi.de</a> )
DEX	Data Exchange, IHE-Profil
DFG	Deutsche Forschungsgemeinschaft ( <a href="http://www.dfg.de">www.dfg.de</a> )
DGAI	Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin e.V. ( <a href="http://www.dgai.de">www.dgai.de</a> )
DH	Digital Health
DICOM	Digital Imaging and Communications in Medicine ( <a href="http://medical.nema.org">http://medical.nema.org</a> )
DICOM4CHEE	Open Source DICOM Toolkit
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information ( <a href="http://www.dimdi.de">www.dimdi.de</a> )
DIN	Deutsches Institut für Normung e.V. ( <a href="http://www.din.de">www.din.de</a> )
DIVI	Deutsche Interdisziplinäre Vereinigung für Intensiv- und Notfallmedizin e.V. ( <a href="http://www.divi-org.de">www.divi-org.de</a> )
DKFZ	Deutsches Krebsforschungszentrum ( <a href="http://www.dkfz.de">www.dkfz.de</a> )
DKTK	Deutsches Konsortium für Translationale Krebsforschung ( <a href="http://www.dkfz.de/de/dkkt">www.dkfz.de/de/dkkt</a> )
DLR	Deutsches Zentrum für Luft- und Raumfahrt, PT ( <a href="http://www.dlr.de">www.dlr.de</a> )
DM	Data Management
DMP	Data Management and Sharing Plan
DMS	Dokumenten Management System
DMSG	Deutsche Multiple Sklerose Gesellschaft, Bundesverband e. V. ( <a href="http://www.dmsg.de">www.dmsg.de</a> )
DNA	Deoxyribonucleic acid (Desoxyribonukleinsäure)
DNEbM	Deutsches Netzwerk Evidenzbasierte Medizin e.V. ( <a href="http://www.ebm-netzwerk.de">www.ebm-netzwerk.de</a> )
DNVF	Deutsches Netzwerk Versorgungsforschung ( <a href="http://www.dnvf.de">www.dnvf.de</a> )

DOI	Digital Object Identifier System der IDF ( <a href="http://www.doi.org">www.doi.org</a> )
DRKS	Deutsches Register Klinischer Studien ( <a href="http://www.drks.de">www.drks.de</a> )
DRYAD	US-Non-Profit-Einrichtung, die ein Data Repository für die zitierfähige und nachnutzbare Bereitstellung von Forschungsdaten betreibt ( <a href="http://datadryad.org">http://datadryad.org</a> )
DSGVO	Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr – Datenschutz-Grundverordnung (Entwurf)
DsiN	Deutschland sicher im Netz ( <a href="http://www.sicher-im-netz.de">www.sicher-im-netz.de</a> )
DSK	Datenschutzkonferenz – Konferenz der Datenschutzbeauftragten des Bundes und der Länder
DTU	Data Transfer Unit (Datentransferstelle)
DVMD	Deutscher Verband Medizinischer Dokumentare ( <a href="http://www.dvmd.de">www.dvmd.de</a> )
DWH	Data Ware House
DZG	Deutsche Zentren der Gesundheitsforschung
DZHK	Deutsches Zentrum für Herz-Kreislauf-Forschung e.V. ( <a href="http://dzhk.de">http://dzhk.de</a> )
DZIF	Deutsches Zentrum für Infektionsforschung e.V. ( <a href="http://www.dzif.de">www.dzif.de</a> )
DZL	Deutsches Zentrum für Lungenforschung ( <a href="http://www.dzg-lungenforschung.de">www.dzg-lungenforschung.de</a> )
e:Med	e:Med – Maßnahmen zur Etablierung der Systemmedizin; Forschungs- und Förderkonzept des BMBF
EATRIS	European Advanced Translational Research Infrastructure in Medicine, ein ESFRI-Projekt ( <a href="http://www.eatris.eu">www.eatris.eu</a> )
EBI	European Bioinformatics Institute, Teil des EMBL ( <a href="http://www.ebi.ac.uk">www.ebi.ac.uk</a> )
EC	European Commission
ECMA	Internationale Organisation zur Standardisierung von Informations- und Kommunikationssystemen. Früher: European Computer Manufacturers Association ( <a href="http://www.ecma-international.org">www.ecma-international.org</a> )
eCRF	electronic Case Report Form
ECRIN	European Clinical Research Infrastructures Network, seit 2007 als ESFRI-Projekt gefördert ( <a href="http://www.eclin.org">www.eclin.org</a> )
ECRIN-IA	ECRIN Integrating Activity; kollaboratives, von 2012 bis 2015 im FP7 gefördertes Projekt
EDC	Electronic Data Capturing
EFI	Expertenkommission Forschung und Innovation ( <a href="http://www.e-fi.de">www.e-fi.de</a> )
EFMI	European Federation for Medical Informatics ( <a href="http://www.efmi.org">www.efmi.org</a> )
EG	Europäische Gemeinschaft
EGA	European Genome-phenome Archive ( <a href="http://www.ebi.ac.uk/ega/home">www.ebi.ac.uk/ega/home</a> )
EHR	Electronic Health Record
EHR4CR	Electronic Health Records for Clinical Research, im Rahmen der IMI gefördertes EU-Projekt ( <a href="http://www.ehr4cr.eu">www.ehr4cr.eu</a> )

EKG	Elektrokardiogramm
EKT	(Greek) National Documentation Center ( <a href="http://www.ekt.gr/en">www.ekt.gr/en</a> )
ELIXIR	European Life Sciences Infrastructure for Biological Information, ein ESFRI-Projekt ( <a href="http://www.elixir-europe.org">www.elixir-europe.org</a> )
ELSA	Förderschwerpunkt des BMBF zu ethischen, rechtlichen und sozialen Aspekten der modernen Lebenswissenschaften und der Biotechnologie
EMA	European Medicines Agency ( <a href="http://www.ema.europa.eu">www.ema.europa.eu</a> )
EMBL	European Molecular Biology Laboratory ( <a href="http://www.embl.org">www.embl.org</a> )
EMR	Electronic Medical Record
EN	Europäische Norm des CEN
ENA	European Nucleotide Archive am EMBL-EBI ( <a href="http://www.ebi.ac.uk/ena">www.ebi.ac.uk/ena</a> )
ENISA	European Union Agency for Network and Information Security ( <a href="http://www.enisa.europa.eu">www.enisa.europa.eu</a> )
EORTC	European Organisation for Research and Treatment of Cancer ( <a href="http://www.eortc.be">www.eortc.be</a> )
EOSC	European Open Science Cloud; Initiative der Europäischen Kommission ( <a href="http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud">http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud</a> )
EP	Europäisches Parlament ( <a href="http://www.europarl.europa.eu">www.europarl.europa.eu</a> )
ePA	elektronische Patientenakte
E-PIX	Im Rahmen des GANI_MED-Projekts entwickelte und mit dem PIX-Profil von IHE kompatible MPI-Software
ePRO	electronic Patient Reported Outcomes
epSOS	European Patients Smart Open Services; EU-Projekt zur Förderung der Interoperabilität von eHealth-Systemen für Notfalldaten und elektronisches Rezept ( <a href="http://www.epsos.eu">www.epsos.eu</a> )
ERIC	European Research Infrastructure Consortium; europäisches Rechtsinstrument für Forschungsinfrastrukturen ( <a href="http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric">http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric</a> )
ESBB	European, Middle Eastern & African Society for Biopreservation and Biobanking; Unterorganisation der ISBER ( <a href="http://www.esbb.org">www.esbb.org</a> )
ESBBperanto	Working Group der ESBB zur Harmonisierung der Daten- und Probenbeschreibungen in Biobanken
eSDI	electronic Source Data Interchange
ESF	European Science Foundation: Zusammenschluss von 77 Forschungsorganisationen aus 30 europäischen Ländern mit Sitz in Straßburg. Deutsche Mitglieder z. B. DFG, MPG und Helmholtz-Gemeinschaft ( <a href="http://www.esf.org">www.esf.org</a> )
ESFRI	European Strategy Forum on Research Infrastructures ( <a href="http://cordis.europa.eu/esfri">http://cordis.europa.eu/esfri</a> )
ESWC	European Semantic Web Conference ( <a href="http://eswc-conferences.org">http://eswc-conferences.org</a> )
ETH	Eidgenössische Technische Hochschule (Zürich)

ETL	Extract, Transform, Load: Kurzform für den Prozess, Daten aus mehreren, heterogenen Datenquellen selektiv zu lesen, zu transformieren und in einer einheitlichen Zielstruktur abzuspeichern
eTMF	electronic Trial Master File
eTRIKS	European Translational Information and Knowledge Management Services ( <a href="http://www.etriks.org">www.etriks.org</a> )
EU-ADR	Exploring and understanding Adverse Drug Reactions by Integrative Mining of Clinical Records and Biomedical Knowledge: In FP7 gefördertes EU-Projekt ( <a href="http://www.alert-project.org">www.alert-project.org</a> )
EU-DSG-VO	EU-Datenschutzgrundverordnung
EU-OPENSREEN	European Infrastructure of Open Screening Platforms for Chemical Biology; als ERIC organisierte Forschungsinfrastruktur ( <a href="http://eu-openscreen.eu">http://eu-openscreen.eu</a> )
EuGH	Gerichtshof der Europäischen Gemeinschaften ( <a href="http://curia.europa.eu">http://curia.europa.eu</a> )
EURAT	Projekt des Marsilius-Kollegs an der Universität Heidelberg zu den Ethischen und Rechtlichen Aspekten der Totalsequenzierung des menschlichen Genoms ( <a href="http://www.uni-heidelberg.de/totalsequenzierung">www.uni-heidelberg.de/totalsequenzierung</a> )
EWR	Europäischer Wirtschaftsraum
FAIR	Findable, Accessible, Interoperable, Reusable
FAIRDOM	Support and Service Network for European Systems Biology; Ziel ist die Unterstützung von Projekten in der Standardisierung, dem Management und der Disseminierung von Daten und Modellen nach dem FAIR-Prinzip: Findable, Accessible, Interoperable, Reusable ( <a href="http://www.fair-dom.org">www.fair-dom.org</a> )
FASTA	Textbasiertes Format zur Speicherung von Sequenzdaten
FASTQ	Textbasiertes Format zur Speicherung von Sequenz- und Qualitätsdaten
FastX Toolkit	Kommandozeilen-Tools zur Vorverarbeitung von FASTA- und FASTQ-Dateien ( <a href="http://hannonlab.cshl.edu/fastx_toolkit">http://hannonlab.cshl.edu/fastx_toolkit</a> )
FDA	US Food and Drug Administration ( <a href="http://www.fda.gov">www.fda.gov</a> )
FHIR	Fast Healthcare Interoperability Resources; HL7-Standard ( <a href="http://hl7.org/fhir">http://hl7.org/fhir</a> )
FKZ	Förderkennzeichen
FMRIB	Oxford Centre for Functional MRI of the Brain, Nuffield Department of Clinical Neurosciences, University of Oxford ( <a href="http://www.fmrib.ox.ac.uk">www.fmrib.ox.ac.uk</a> )
FP	Framework Programme for Research and Technology Development: Hauptinstrument der EU zur Förderung von Forschung und Entwicklung
FP7	7. FP der EU
FR	Federation representative
FreeSurfer	Open-Source-Software zur Verarbeitung und Analyse von MRI-Daten des (menschlichen) Gehirns ( <a href="http://freesurfer.net">http://freesurfer.net</a> )
FSL	FMRIB Software Library ( <a href="http://fsl.fmrib.ox.ac.uk">http://fsl.fmrib.ox.ac.uk</a> )

FTP	File Transfer Protocol: Standardprotokoll zur Übertragung von Dateien im Internet
GANI_MED	Greifswald Approach to Individualized Medicine ( <a href="http://www.medizin.uni-greifswald.de/gani_med">www.medizin.uni-greifswald.de/gani_med</a> )
GATK	Genome Analysis Toolkit ( <a href="http://www.broadinstitute.org/gatk">www.broadinstitute.org/gatk</a> )
GB	Gigabyte(s)
G-BA	Gemeinsamer Bundesausschuss, Gremium der gemeinsamen Selbstverwaltung von Ärzten, Krankenkassen und Krankenhäusern ( <a href="http://www.g-ba.de">www.g-ba.de</a> )
GBA	German Biobank Alliance
GBN	German Biobank Node; deutscher nationaler Hub im BBMRI ERIC
GCP	Good Clinical Practice, Regelwerk der ICH
GCS	Glasgow Coma Scale ( <a href="http://glasgowcomascale.org">http://glasgowcomascale.org</a> )
GEKID	Gesellschaft epidemiologischer Krebsregister in Deutschland e.V. ( <a href="http://www.gekid.de">www.gekid.de</a> )
Gene Ontology	Projekt zur Erstellung eines kontrollierten Vokabulars für genetische Daten ( <a href="http://www.geneontology.org">www.geneontology.org</a> )
GenePattern	Open-Source-Analysewerkzeug für genetische Daten ( <a href="http://www.broadinstitute.org/cancer/software/genepattern">www.broadinstitute.org/cancer/software/genepattern</a> )
GeneXplain	Portal mit Werkzeugen und Workflow-Management für Bioinformatik, Systembiologie und Cheminformatik ( <a href="http://genexplain.com">http://genexplain.com</a> )
GenoPerspektiv	Zum Umgang mit genomischen Hochdurchsatzdaten: Die Perspektiven von Klinik, Ethik, Recht und biomedizinischer Informationstechnologie; gefördertes BMBF-Projekt ( <a href="http://www.genoperspektiv.de">www.genoperspektiv.de</a> )
GEO	Gene Expression Omnibus ( <a href="http://www.ncbi.nlm.nih.gov/geo">www.ncbi.nlm.nih.gov/geo</a> )
GFZ	Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum ( <a href="http://www.gfz-potsdam.de">www.gfz-potsdam.de</a> )
GG	Grundgesetz der Bundesrepublik Deutschland
GGBN	Global Genome Biodiversity Network ( <a href="http://www.ggbn.org">http://www.ggbn.org</a> )
GI	Gesellschaft für Informatik e.V. ( <a href="http://www.gi-ev.de">www.gi-ev.de</a> )
gICS	generic Informed Consent Service der Universitätsmedizin Greifswald
GKV	Gesetzliche Krankenversicherung
GKV-Spitzenverband	Spitzenverband Bund der Krankenkassen gemäß § 217a SGB V ( <a href="http://www.gkv-spitzenverband.de">www.gkv-spitzenverband.de</a> )
GMDS	Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. ( <a href="http://www.gmds.de">www.gmds.de</a> )
GMS	German Medical Science: Webportal der German Medical Science gGmbH ( <a href="http://www.egms.de">www.egms.de</a> )
GNU	Rekursives Akronym für "GNU's Not Unix": 1984 gegründetes Projekt zur Erstellung eines freien (Open Source) Unix Betriebssystems ( <a href="http://www.gnu.org">www.gnu.org</a> )

GoÄ	Gebührenordnung für Ärzte
gPAS	generic Pseudonym Administration Service
GPS	Global Positioning System
GPU	Graphics Processing Unit
GSP	Good Scientific Practice
GTDS	Gießener Tumordokumentations-System ( <a href="http://www.gtds.de">www.gtds.de</a> )
GWAS	Genomweite Assoziationsstudie(n)
GWK	Gemeinsame Wissenschaftskonferenz ( <a href="http://www.gwk-bonn.de">www.gwk-bonn.de</a> )
GZIP	GNU ZIP, freies Kompressionsprogramm
HaPI	Health and Psychosocial Instruments ( <a href="http://www.bmdshapi.com">www.bmdshapi.com</a> )
HBFG	Hochschulbauförderungsgesetz
HDOT	Health Data Ontology Trunk
HEGP	Hôpital Européen Georges-Pompidou ( <a href="http://www.hegp.fr">www.hegp.fr</a> )
HGF	Helmholtz-Gemeinschaft Deutscher Forschungszentren ( <a href="http://www.helmholtz.de">www.helmholtz.de</a> )
HIPAA	US Health Insurance Portability and Accountability Act
HIS	Health Information System
HIV	Human Immunodeficiency Virus
HL7	Health Level Seven; Internationale SDO für den Bereich der Interoperabilität von IT-Systemen im Gesundheitswesen ( <a href="http://www.hl7.org">www.hl7.org</a> )
Horizon 2020	The EU Framework Programme for Research and Innovation ( <a href="http://ec.europa.eu/programmes/horizon2020">http://ec.europa.eu/programmes/horizon2020</a> )
HPLC	High-Performance Liquid Chromatography
HRK	Hochschulrektorenkonferenz ( <a href="http://www.hrk.de">www.hrk.de</a> )
HTTP	Hyper Text Transfer Protocol
i.s.h.med	In SAP for Healthcare integriertes Klinisches Informationssystem der Siemens AG
i2b2	Informatics for Integrating Biology and the Bedside ( <a href="http://www.i2b2.org">www.i2b2.org</a> )
I2B2	Informatics for Integrating Biology and the Bedside ( <a href="http://www.i2b2.org">www.i2b2.org</a> )
IaaS	Infrastructure as a Service
ICD	International Statistical Classification of Diseases and Related Health Problems
ICGC	International Cancer Genome Consortium ( <a href="http://icgc.org">http://icgc.org</a> )
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use ( <a href="http://www.ich.org">www.ich.org</a> )
ICMJE	International Committee of Medical Journal Editors ( <a href="http://www.icmje.org">www.icmje.org</a> )
ICR	Intelligent Character Recognition; automatisierte Erkennung eines Zeichenmusters als Buchstabe oder Zahl; siehe auch OCR
ICT	Information Communication Technology
ID	Identifikationsnummer

IDF	International DOI Foundation ( <a href="http://www.doi.org">www.doi.org</a> )
IDRT	Integrated Data Repository Toolkit: TMF-Projekt zur Erarbeitung von Instrumenten und Methoden zur Integration verteilter und heterogener Datenbestände für die klinische und translationale Forschung ( <a href="http://www.tmf-ev.de/idrt">www.tmf-ev.de/idrt</a> )
IEC	International Electrotechnical Commission ( <a href="http://www.iec.ch">www.iec.ch</a> )
IEEE	International non-profit organization and professional association for the advancement of technology, ursprgl.: Institute of Electrical and Electronics Engineers ( <a href="http://www.ieee.org">www.ieee.org</a> )
IFB	Integrierte Forschungs- und Behandlungszentren; Fördermaßnahme des BMBF
IGES	Institut für Gesundheits- und Sozialforschung GmbH ( <a href="http://www.iges.de">www.iges.de</a> )
IHE	Integrating the Healthcare Enterprise ( <a href="http://www.ihe.net">www.ihe.net</a> )
IIT	Investigator initiated trial
IKT	Informations- und Kommunikationstechnologie
ImageJ	Image Processing and Analysis in Java ( <a href="http://imagej.nih.gov/ij">http://imagej.nih.gov/ij</a> )
IMI	Innovative Medicines Initiative ( <a href="http://www.imi-europe.org">www.imi-europe.org</a> )
IMIA	International Medical Informatics Association ( <a href="http://www.imia.org">www.imia.org</a> )
IMISE	Institut für Medizinische Informatik, Statistik und Epidemiologie der Universität Leipzig ( <a href="http://www.imise.uni-leipzig.de">www.imise.uni-leipzig.de</a> )
INCF	International Neuroinformatics Coordination Facility ( <a href="http://www.incf.org">www.incf.org</a> )
Infrafrontier	ESFRI-Projekt zur Infrastruktur für Mausmodellforschung ( <a href="http://www.infrafrontier.eu">www.infrafrontier.eu</a> )
IntAct	Molekulare Interaktions-Datenbank ( <a href="http://www.ebi.ac.uk/intact">www.ebi.ac.uk/intact</a> )
iOS	Operating System für Apple iPhones und iPads
IP	Internet Protocol
IPL	In-Vivo-Pathophysiologie-Labor der Helmholtz-Gemeinschaft
IQSS	The Institute for Quantitative Social Science der Harvard University ( <a href="http://www.iq.harvard.edu">www.iq.harvard.edu</a> )
iRODS	Integrated Rule-Oriented Data System; Open-Source-Software zum Datenmanagement in Forschungseinrichtungen und öffentlicher Verwaltung ( <a href="http://irods.org">http://irods.org</a> )
ISBER	International Society for Biological and Environmental Repositories ( <a href="http://www.isber.org">www.isber.org</a> )
ISBN	International Standard Book Number: Eindeutige Nummer zur Kennzeichnung von Büchern und anderen selbstständigen, nicht periodischen Veröffentlichungen ( <a href="http://www.isbn-international.org">www.isbn-international.org</a> )
ISO	International Organization for Standardization ( <a href="http://www.iso.org">www.iso.org</a> )
ISO/TS	ISO Technical Specification

ISPOR	International Society for Pharmacoeconomics and Outcomes Research ( <a href="http://www.ispor.org">www.ispor.org</a> )
ITK	Insight Segmentation and Registration Toolkit der US National Library of Medicine ( <a href="http://www.itk.org">www.itk.org</a> )
IT-QM	IT-Infrastruktur und Qualitätsmanagement (AG)
JavaScript	Einfache und teilweise von der ECMA standardisierte Programmiersprache zur Ausführung von Programmlogik in Webbrowsern
JIC	Joint Initiative Council: Joint Initiative on SDO Global Health Informatics Standardization ( <a href="http://www.jointinitiativecouncil.org">www.jointinitiativecouncil.org</a> )
JMIR	Journal of Medical Internet Research ( <a href="http://www.jmir.org">www.jmir.org</a> )
JPEG	Komprimiertes Bilddateiformat der Joint Photographic Experts Group ( <a href="http://www.jpeg.org">www.jpeg.org</a> )
JSON	JavaScript Object Notation: Kompaktes Datenformat für die Datenübertragung auf Basis der JavaScript-Syntax
JSON-LD	JSON für Linked Data
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
k-Anonymität	Eine Datensammlung ist k-anonym, wenn jede Merkmalskombination, die potentiell für einen reidentifizierenden Abgleich genutzt werden könnte, in mindestens k Datensätzen vorkommt
KAS	Klinisches Arbeitsplatz-System
KBV	Kassenärztliche Bundesvereinigung ( <a href="http://www.kbv.de">www.kbv.de</a> )
KfR	Kommission für IT-Infrastruktur der DFG (früher: Kommission für Rechenanlagen)
KFRG	Gesetz zur Weiterentwicklung der Krebsfrüherkennung und zur Qualitätssicherung durch klinische Krebsregister – Krebsfrüherkennungs- und -registergesetz
KHEntgG	Gesetz über die Entgelte für voll- und teilstationäre Krankenhausleistungen – Krankenhausentgeltgesetz
KIS	Krankenhausinformationssystem
KISREK	BMBF-gefördertes Projekt zur KIS-basierten Unterstützung der Patientenrekrutierung in klinischen Studien ( <a href="http://www.tmf-ev.de/kis-rekrutierung">www.tmf-ev.de/kis-rekrutierung</a> )
KKNMS	Krankheitsbezogenes Kompetenznetz Multiple Sklerose ( <a href="http://www.kompetenznetz-multiplesklerose.de">www.kompetenznetz-multiplesklerose.de</a> )
KKS	Koordinierungszentrum für Klinische Studien ( <a href="http://www.kks-netzwerk.de">www.kks-netzwerk.de</a> )
KKS-Netzwerk	Netzwerk der Koordinierungszentren für Klinische Studien ( <a href="http://www.kks-netzwerk.de">www.kks-netzwerk.de</a> )
KN	Kompetenznetz ( <a href="http://www.kompetenznetze-medizin.de">www.kompetenznetze-medizin.de</a> )

KNIME	KoNstanz Information MinEr ( <a href="http://www.knime.org">www.knime.org</a> )
KOM	Kommunikation mit Studienzentren
KoRegIT	TMF-Projekt zur Erstellung eines Anforderungskatalogs zur IT-Unterstützung von Kohorten und Registern
LabID <sub>tr</sub>	kryptographisch transformierte Proben-ID (Labor-ID)
LABIMI/F	Langzeitarchivierung biomedizinischer Forschungsdaten; von der DFG gefördertes Projekt ( <a href="http://www.labimi-f.med.uni-goettingen.de">www.labimi-f.med.uni-goettingen.de</a> )
LC	Liquid Chromatography
l-Diversität	Maß für die Unterschiedlichkeit sensitiver Attribute in k-anonymisierten Datenbanken: In jedem Block mit k identischen quasi-identifizierenden Daten gibt es mindestens l unterschiedliche Ausprägungen sensitiver Attribute
LE	Leistungserbringer
LexGrid	Lexical Grid: Von der Mayo Clinic koordiniertes Projekt zur Standardisierung von Schnittstellen und Tools auf Basis eines gemeinsamen terminologischen Modells zur Vereinfachung der Nutzung bestehender Terminologien und Klassifikationen ( <a href="http://informatics.mayo.edu/LexGrid">http://informatics.mayo.edu/LexGrid</a> )
LGPL	GNU Lesser General Public License ( <a href="http://www.gnu.org/licenses/lgpl.html">www.gnu.org/licenses/lgpl.html</a> )
LIBE	Committee on Civil Liberties, Justice and Home Affairs - Ausschuss für bürgerliche Freiheiten, Justiz und Inneres des EP ( <a href="http://www.europarl.europa.eu/committees/de/libe/home.html">www.europarl.europa.eu/committees/de/libe/home.html</a> )
LIFE	Leipziger Interdisziplinärer Forschungskomplex zu molekularen Ursachen umwelt- und lebensstilassoziierter Erkrankungen ( <a href="http://www.uni-leipzig-life.de">www.uni-leipzig-life.de</a> )
LIMS	Laboratory Information Management System
LIS	Laborinformationssystem
LLC	Limited Liability Company
LOC	(REDCap) Library Oversight Committee
LOINC	Logical Observation Identifiers Names and Codes ( <a href="http://www.loinc.org">www.loinc.org</a> )
LORIS	Longitudinal Online Research and Imaging System; webbasierte Open-Source-Software für Daten- und Projektmanagement in Neuromaging-Studien ( <a href="http://mcin-cnim.ca/neuroimagingtechnologies/loris">http://mcin-cnim.ca/neuroimagingtechnologies/loris</a> )
MAGIC	Mainzelliste, Samply.Auth und der Generische Informed Consent Service als Open-Source-Werkzeuge für Identitäts-, Einwilligungs- und Rechtemanagement in der medizinischen Verbundforschung; DFG-gefördertes Verbundprojekt
Mainzelliste	Webbasierte Open-Source Pseudonymisierungsplattform der Universitätsmedizin Mainz ( <a href="http://www.unimedizin-mainz.de/imbei/informatik/opensource/mainzelliste.html">http://www.unimedizin-mainz.de/imbei/informatik/opensource/mainzelliste.html</a> )
MAKS	Makros zur Auswertung Klinischer Studien: Im Rahmen eines TMF-Projekts erstellte und validierte Bibliothek von SAS-Makros zur Auswertung von Studiendaten in der CDISC SDTM Struktur

MB	Megabyte(s)
MD	Muskeldystrophie
MDC	Max-Delbrück-Centrum für Molekulare Medizin ( <a href="http://www.mdc-berlin.de">www.mdc-berlin.de</a> )
mdi	mdi – Forum der Medizin_Dokumentation und Medizin_Informatik; von den beiden Verbänden BVMI und DVMD gemeinsam herausgegebene Fachzeitschrift
MDPE	Medical Data and Picture Exchange: Im KN POH entwickelte Teleradiologielösung auf Basis des Modells A der generischen Datenschutzkonzepte der TMF.
MDR	EU Medical Devices Regulation
MDR	Metadata Repository
MEDDEV	Medical Devices Guidance Document der EU
MedDRA	Medical Dictionary for Regulatory Activities ( <a href="http://www.meddransso.com">www.meddransso.com</a> )
medico	KIS der Siemens AG
MERIL	Mapping of the European Research Infrastructure Landscape ( <a href="http://portal.meril.eu">http://portal.meril.eu</a> )
MetaVelvet	Metagenom Assembler ( <a href="http://metavelvet.dna.bio.keio.ac.jp">http://metavelvet.dna.bio.keio.ac.jp</a> )
METeOR	Australian Metadata Online Registry ( <a href="http://meteor.aihw.gov.au">http://meteor.aihw.gov.au</a> )
MethInfraNet	Maßnahmen zur methodischen und infrastrukturellen Vernetzung für Qualitäts- und Effizienzsteigerung in der medizinischen Forschung; BMBF-Zuwendung für Ausbau und Verstärkung der TMF
MFT	Medizinischer Fakultätentag ( <a href="http://www.mft-online.de">www.mft-online.de</a> )
MHH	Medizinische Hochschule Hannover ( <a href="http://www.mh-hannover.de">www.mh-hannover.de</a> )
MIABIS	Minimum Information About BIobank data Sharing (Standard)
MIAME	Minimum Information About a Microarray Experiment (Standard)
MIAPE	Minimum Information About a Proteomics Experiment (Standard)
MIE	Medical Informatics Europe: Regelmäßige Konferenz der EFMI
MIGS	Minimum Information about a (Meta)Genome Sequence (Standard)
MInCa	Mobile Information Capture; EDC-System für mobile Datenerfassung und spätere Synchronisierung mit einem zentralen Datenmanagement
MIPAV	Medical Image Processing, Analysis, and Visualization ( <a href="http://mipav.cit.nih.gov">http://mipav.cit.nih.gov</a> )
MIRC	Medical Imaging Resource Center, von der RSNA initiiertes Open Source Software-Projekt ( <a href="https://rsna.org/MIRC.aspx">https://rsna.org/MIRC.aspx</a> )
MITK	The Medical Imaging Interaction Toolkit ( <a href="http://mitk.org">http://mitk.org</a> )
MO Forum	ESF Member Organisation Forum on Research Infrastructures
MoCoMed	Mobiles Computing in der Medizin; gemeinsame Arbeitsgruppe von GMDS und GI ( <a href="http://www.mocomed.org">www.mocomed.org</a> )

MOSAIC	Modular Systematic Approach to Implement a Centralized Data Management – Open-Source-Werkzeuge für zentrales Datenmanagement in der epidemiologischen Forschung; DFG-gefördertes Projekt des Instituts für Community Medicine der Universitätsmedizin Greifswald ( <a href="https://mosaic-greifswald.de">https://mosaic-greifswald.de</a> )
mosaicQA	R-Package zur Erstellung von Qualitätsberichten für metrische und kategoriale Daten
MPG	Gesetz über Medizinprodukte - Medizinproduktegesetz
MPG	Max-Planck-Gesellschaft ( <a href="http://www.mpg.de">www.mpg.de</a> )
MPI	Master Patient Index
MR	Magnetic Resonance
MRC	UK Medical Research Council ( <a href="http://www.mrc.ac.uk">www.mrc.ac.uk</a> )
MRCT	Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard ( <a href="http://mrcctcenter.org">http://mrcctcenter.org</a> )
MRI	Magnetic Resonance Imaging
mRNA	messenger-RNA
MRT	Magnetresonanztomographie
MS	Multiple Sklerose
MSI	Metabolomics Standards Initiative ( <a href="http://www.metabolomics-msi.org">www.metabolomics-msi.org</a> )
MVC	Master Value Catalogue
NA	Normenausschuss des DIN
NAKO	NAKO Gesundheitsstudie, bundesweite Gesundheitsstudie mit 200.000 Teilnehmern ( <a href="http://www.nako.de">www.nako.de</a> )
NCI	US National Cancer Institute ( <a href="http://www.cancer.gov">www.cancer.gov</a> )
NCT	Nationales Centrum für Tumorerkrankungen Heidelberg ( <a href="http://www.dkfz.de/de/nct">www.dkfz.de/de/nct</a> )
NEMA	US National Electrical Manufacturers Association ( <a href="http://www.nema.org">www.nema.org</a> )
NER	Nationale Ethikrat ( <a href="http://www.ethikrat.org">www.ethikrat.org</a> )
NFDI	Nationale Forschungsdateninfrastruktur; Empfehlung des RfII
NGS	Next-Generation Sequencing
NI-DM	Neuroimaging Data Model der NIDASH
NIDA	US National Institute on Drug Abuse ( <a href="http://www.drugabuse.gov">www.drugabuse.gov</a> )
NIDA Data Share	Data Share Website des NIDA ( <a href="https://datashare.nida.nih.gov">https://datashare.nida.nih.gov</a> )
NIDASH	Neuroimaging Data Sharing Task Force der INCF
NI-DM	Neuroimaging Data Model der NIDASH
NIFTI	Neuroimaging Informatics Technology Initiative ( <a href="http://nifti.nimh.nih.gov">http://nifti.nimh.nih.gov</a> )
NIH	US National Institutes of Health ( <a href="http://www.nih.gov">www.nih.gov</a> )
NIRK	Network for Ichthyoses and Related Keratinization Disorders ( <a href="http://www.netzwerk-ichthyose.de">www.netzwerk-ichthyose.de</a> )
NIST	US National Institute of Standards and Technology ( <a href="http://www.nist.gov">www.nist.gov</a> )
NLP	Natural Language Processing

NoSQL	Not Only SQL; Datenbanktechnologie, die die Beschränkung von RDBMS auf die Verwaltung rein tabellarischer Daten aufhebt
NRW	Nordrhein-Westfalen
NSA	US National Security Agency ( <a href="http://www.nsa.gov">www.nsa.gov</a> )
OASIS	Organization for the Advancement of Structured Information Standards ( <a href="http://www.oasis-open.org">www.oasis-open.org</a> )
OAuth	offenes Autorisierungsprotokoll für verteilte Anwendungen
OBO	Open Biomedical Ontologies (Foundry) ( <a href="http://www.obofoundry.org">www.obofoundry.org</a> )
OCR	Optical Character Recognition; automatisierte Erkennung eines Zeichenschemas als Buchstabe oder Zahl; siehe auch ICR
ODM	Operational Data Model (CDISC-Standard)
OECD	Organisation for Economic Co-operation and Development ( <a href="http://www.oecd.org">www.oecd.org</a> )
OFFIS	OFFIS e.V., ausgegründetes Institut für Informatik der Universität Oldenburg und der Jade Hochschule ( <a href="http://www.offis.de">www.offis.de</a> )
OG	Offene Gesellschaft, Rechtsform in Österreich
OHDSI	Observational Health Data Sciences and Informatics; Initiative zur Nutzung von Gesundheitsdaten für large scale analytics ( <a href="http://www.ohdsi.org">www.ohdsi.org</a> )
OME	Open Microscopy Environment ( <a href="http://www.openmicroscopy.org">www.openmicroscopy.org</a> )
OMERO	OME Remote Objects; Client-Server-Software für das Visualisieren, Managen und Annotieren wissenschaftlicher Bilddaten
OMICs	Suffix zur Kennzeichnung eines Teilgebiets der molekularen Biologie
OMOP	Observational Medical Outcomes Partnership: Historische Initiative aus den USA zur Entwicklung von Methoden und Verfahren zur Nutzung klinischer Beobachtungsdaten zur Überwachung der Sicherheit medizinischer Produkte, heute überwiegend aufgegangen in der OHDSI ( <a href="http://omop.org">http://omop.org</a> )
OP	Operation, Operationsaal
openBIS	offenes, verteiltes System für das Management biologischer Daten ( <a href="http://www.cisd.ethz.ch/software/openBIS">www.cisd.ethz.ch/software/openBIS</a> )
OpenClinica	Open-Source-Software für klinische Studien mit Studiendatenmanagement- und EDC-Funktionen ( <a href="http://www.openclinica.org">www.openclinica.org</a> )
openEHR	Internationale Organisation zur Entwicklung interoperabler Electronic Healthcare Records ( <a href="http://www.openehr.org">www.openehr.org</a> )
openSEEK	Plattform zum Austausch von Daten und Modellen in der Systembiologie (siehe FAIRDOM und openBIS)
OpenSpecimen	auf der Basis von caTISSUE weiterentwickelte und quelloffene Software zur Probenverwaltung ( <a href="http://www.openspecimen.org">www.openspecimen.org</a> )
OpenStack	Open-Source-Software für Clouds ( <a href="http://www.openstack.org">www.openstack.org</a> )
OpenTrials	Initiative und Web-Plattform zur Sammlung und Verlinkung aller verfügbaren Informationen zu allen klinischen Studien ( <a href="http://opentrials.net">http://opentrials.net</a> )

OPS	Operationen- und Prozedurenschlüssel; vom DIMDI herausgegebener Katalog zur Verschlüsselung medizinischer Prozeduren im Krankenhaus und ambulanter Operationen
Orphanet	Referenz-Portal für Informationen über seltene Krankheiten und Orphan Drugs ( <a href="http://www.orpha.net">www.orpha.net</a> )
OSSE	Open-Source-Registersystem für Seltene Erkrankungen in der EU; vom BMG gefördertes Projekt im Rahmen des Aktionsplans für Menschen mit seltenen Erkrankungen
OWL	Web Ontology Language
PaaS	Platform as a Service
PACS	Picture Archiving and Communication System
PB	Petabyte
PCORI	US Patient-Centered Outcomes Research Institute ( <a href="http://www.pcori.org">www.pcori.org</a> )
PCORnet	US National Patient-Centered Clinical Research Network; Initiative des PCORI ( <a href="http://www.pcornet.org">www.pcornet.org</a> )
PCR	Polymerase-Chain-Reaction (Polymerasekettenreaktion)
PC	Personal Computer
PDA	Personal Digital Assistant
PDB	Protein Data Bank der RCSB ( <a href="http://www.rcsb.org/pdb">www.rcsb.org/pdb</a> )
PDF	Portable Document Format von Adobe ( <a href="http://www.adobe.com">www.adobe.com</a> )
PDQ	Patients Demographics Query, IHE-Profil zur Abfrage demographischer Daten zu Patienten in verteilten Systemen
PDV	Patientendatenverwaltung
PET	Positronenemissionstomographie
PhEMA	Phenotype Execution Modeling Architecture ( <a href="http://informatics.mayo.edu/phema">http://informatics.mayo.edu/phema</a> )
PhUSE	Pharmaceutical Users Software Exchange ( <a href="http://www.phuse.info">www.phuse.info</a> )
PID	Patientenidentifikator
PIX	Patient Identifier Cross-referencing, IHE-Profil zum domänenübergreifenden Abgleich von Patienten-Identifikatoren
PLoS	Public Library of Science ( <a href="http://www.plos.org">www.plos.org</a> )
PM	Probandenmanagement
p-medicine	From data sharing and integration via VPH models to personalized medicine; Von der EU kofinanziertes Forschungsprojekt ( <a href="http://www.p-medicine.eu">www.p-medicine.eu</a> )
PNG	Portable Network Graphics, komprimiertes Format für Rastergrafiken
POH	KN Pädiatrische Onkologie und Hämatologie ( <a href="http://www.kompetenznetz-paed-onkologie.de">www.kompetenznetz-paed-onkologie.de</a> )
PPP	Public Private Partnership
PRIDE	Proteomics Identifications database ( <a href="http://www.ebi.ac.uk/pride">www.ebi.ac.uk/pride</a> )

PRO	Patient-Reported Outcome
ProMISE	Project Manager Internet Server; am Leiden University Medical Center entwickeltes, webbasiertes CDMS ( <a href="http://www.msbi.nl/promise">http://www.msbi.nl/promise</a> )
PROV	Provenance; Kürzel und Namespace-Bezeichner des W3C für Dokumente zur Standardisierung von Herkunftsinformationen
PT	Projektträger
PubMed	Online Datenbank der U.S. National Library of Medicine ( <a href="http://pubmed.gov">http://pubmed.gov</a> )
QIIME	Quantitative Insights Into Microbial Ecology ( <a href="https://qiime.org">https://qiime.org</a> )
QM	Qualitätsmanagement
QRPH	IHE Quality, Research and Public Health Domain
QuaSi-Niere	Qualitätssicherung in der Nierenersatztherapie – QuaSi-Niere gGmbH ( <a href="http://www.quasi-niere.de">www.quasi-niere.de</a> )
QUEXML	XML-basierter Abfragestandard der DDI
R	Open-Source-Software für statistische Auswertungen ( <a href="http://www.r-project.org">www.r-project.org</a> )
RCSB	Research Collaboratory for Structural Bioinformatics; Zusammenschluss von Forschungseinrichtungen im Bereich der Bioinformatik ( <a href="http://home.rcsb.org">http://home.rcsb.org</a> )
RDA	Research Data Alliance ( <a href="http://europe.rd-alliance.org">http://europe.rd-alliance.org</a> )
RDBMS	Relationales Datenbank Management System
RDF	Resource Description Framework: Formale Sprache zur Bereitstellung von Metadaten im WWW ( <a href="http://www.w3.org/RDF">www.w3.org/RDF</a> )
RDP	Remote Desktop Protocol, von Microsoft entwickeltes Protokoll, welches Bildschirmaus- und Tastatureingaben über das Netzwerk zwischen entfernten Computern übermittelt.
RDR	Rare Disease Registry
REDCap	Research Electronic Data Capture; Softwareplattform für die Erstellung und Durchführung von Online-Umfragen im wissenschaftlichen Umfeld ( <a href="http://project-redcap.org">http://project-redcap.org</a> )
RELMA	Regenstrief LOINC Mapping Assistant
RENCI	Renaissance Computing Institute an der University of North Carolina in Chapel Hill ( <a href="http://renci.org">http://renci.org</a> )
REST	Representational State Transfer; Standard zur einfachen Datenübertragung in Webanwendungen auf Basis des HTTP-Protokolls
REUSE	Retrieving EHR Useful data for Secondary Exploitation; EU-Projekt
RfII	Rat für Informationsinfrastrukturen ( <a href="http://www.rfii.de">www.rfii.de</a> )
RI	Reidentifizierung
RIF	Rule Interchange Format; Standard des W3C
RIsources	Informationsportal der DFG zu Research Infrastructures ( <a href="http://risources.dfg.de">http://risources.dfg.de</a> )
RNA	Ribonukleinsäure

RöV	Verordnung über den Schutz vor Schäden durch Röntgenstrahlen – Röntgenverordnung
RReMIT	Repository of Registered Migraine Trials; Studienregister der ACTION ( <a href="http://www.action.org/rremit">www.action.org/rremit</a> )
RSNA	Radiological Society of North America ( <a href="http://www.rsna.org">www.rsna.org</a> )
RTPLAN	Radiotherapy Plan; DICOM Information Object Definition
RTSTRUCT	Radiotherapy Structure Set; DICOM Information Object Definition
RWTH	Rheinisch-Westfälische Technische Hochschule Aachen ( <a href="http://www.rwth.de">www.rwth.de</a> )
RZ	Rechenzentrum
SaaS	Software as a Service
SAE	Serious Adverse Event, schwerwiegendes unerwünschtes Ereignis im Rahmen einer Arzneimittelprüfung
SALUS Project	Scalable, Standard based Interoperability Framework for Sustainable Proactive Post Market Safety Studies ( <a href="http://www.salusproject.eu">www.salusproject.eu</a> )
SAM	Sequence Alignment/Map, generisches Format zur Speicherung von Nukleotid-Sequenzen
SCAAR	Swedish Coronary Angiography and Angioplasty Register
SCRUM	Vorgehensmodell der agilen Softwareentwicklung
SCTO	Swiss Clinical Trial Organisation ( <a href="http://www.scto.ch">www.scto.ch</a> )
SDMX	Statistical Data and Metadata Exchange; ISO/IEC 17369
SDO	Standards Development Organization
SDTM	Study Data Tabulation Model (CDISC-Standard)
SecuTrial	professionelle, vollständig browserbasierte Lösung zum Erfassen von Patientendaten in klinischen oder nicht-interventionellen Studien und Patientenregistern ( <a href="http://www.secutrial.com/">www.secutrial.com/</a> )
SEEK	siehe openSEEK
SemanticMDR	im Rahmen von SALUS Project entwickeltes Metadata Repository
SEND	Standards for the Exchange of Non-clinical Data (CDISC-Standard)
SEQ	Sequenz/Sequenzierung
SF-36	Standardisierter Gesundheitsfragebogen mit 36 Items ( <a href="http://www.sf-36.org">www.sf-36.org</a> )
SGB	Sozialgesetzbuch
SH	Schleswig-Holstein
SHARE	CDISC Shared Health and Research Electronic Library
SHARE	Survey of Health, Ageing and Retirement in Europe ( <a href="http://www.share-project.org">www.share-project.org</a> )
SHARPn	von der Mayo Clinic geleitetes Teilprojekt der SHARP-Initiative zum Secondary Use elektronischer Krankenakten (EHR)
SHIP	Study of Health in Pomerania: Vom BMBF geförderte Bevölkerungsgesundheitsstudie ( <a href="http://ship.community-medicine.de">http://ship.community-medicine.de</a> )
SHRINE	Shared Health Research Information Network; Open Source Software für föderative Abfragen auf verteilten I2B2-Installationen

SIGKDD	Special Interest Group on Knowledge Discovery and Data Mining der ACM ( <a href="http://www.kdd.org">www.kdd.org</a> )
SMT	Study Management Tool des Clinical Trial Centers Aachen
SNOMED CT	Systematized Nomenclature of Medicine – Clinical Terms ( <a href="http://www.ihstso.org">www.ihstso.org</a> )
SNP	Single Nucleotide Polymorphism, Variationen von einzelnen Basenpaaren in einem DNA-Strang
SNV	Single Nucleotide Variant
SOAP	Short Oligonucleotide Analysis Package; Alignmentwerkzeug zur Verarbeitung genetischer Daten ( <a href="http://soap.genomics.org.cn">http://soap.genomics.org.cn</a> )
SOAP	Simple Object Access Protocol; vom W3C empfohlener, XML-basierter Protokoll-Standard zur Kommunikation strukturierter Daten mit Webservices per HTTP
Soarian	Workflowmanagementsystem der Siemens AG zur Steuerung der klinischen und administrativen Arbeitsabläufe im Krankenhaus
SOP	Service-Object Pair (Class): Kombination eines DICOM Message Service Element und einer Information Object Definition im DICOM-Standard
SOP	Standard Operating Procedure
SPARQL	rekursives Akronym für SPARQL Protocol And RDF Query Language; vom W3C entwickelte Standard-Abfragesprache
SPM	Statistical Parametric Mapping; Open-Source-Software auf der Basis von MATLAB zur Analyse von Bilddaten aus der funktionellen Bildgebung
SPREC	Standard PREanalytical Code der Biospecimen Science Working Group der ISBER
SPRINT	Simple Parallel R INTERface; R-Package
SPSS	Datenanalysesoftware der Firma SPSS inc. ( <a href="http://www.spss.com">www.spss.com</a> )
SRA	Sequence Read Archive
ST	Segment des EKG
StGB	Strafgesetzbuch
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins; Datenbank für Proteininteraktion ( <a href="http://string-db.org">http://string-db.org</a> )
SWABIK	Software-Werkzeuge für den Austausch von Bilddatenträgern in der klinischen Forschung; Im Rahmen der Ausschreibung zu Instrumenten- und Methodenentwicklungen für die patientenorientierte medizinische Forschung vom BMBF gefördertes Projekt
sysINFLAME	Netzwerk für Systemmedizin chronisch-entzündlicher Erkrankungen; vom BMBF im Rahmen der e:Med-Förderinitiative gefördertes Verbundvorhaben
TA	Therapeutic Area
TA	Technikfolgen-Abschätzung
TAB	Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag ( <a href="http://www.tab.fzk.de">www.tab.fzk.de</a> )

TASTE	Thrombus Aspiration in ST-Elevation myocardial infarction in Scandinavia; multicenter, prospective, randomized, controlled clinical registry trial
TC	Technical Committee (ISO)
TCDP	Trusted Cloud-Datenschutzprofil
TCE	Teaching File and Clinical Trial Export Profile, IHE-Profil
TCGA	The Cancer Genome Atlas ( <a href="http://cancergenome.nih.gov">http://cancergenome.nih.gov</a> )
TCPA	The Cancer Proteome Atlas ( <a href="http://tcpaportal.org/tcpa">http://tcpaportal.org/tcpa</a> )
TFS	MIRC Teaching File System der RSNA ( <a href="http://www.rsna.org/tfs.aspx">www.rsna.org/tfs.aspx</a> )
TIFF	Tagged Image File Format: Dateiformat für Rastergrafiken mit Tags (Auszeichnungen)
TMF	TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. ( <a href="http://www.tmf-ev.de">www.tmf-ev.de</a> )
TNM	Klassifikationssystem der UICC für maligne Tumoren mit den Kategorien (T)umor, (N)odes (Lymphknotenbeteiligung) und (M)etastasen
ToolPool	ToolPool Gesundheitsforschung; von der TMF betriebenes Web-Portal zur Bereitstellung von und Informierung zu Unterstützungsangeboten für IT-Infrastrukturen in der medizinischen Forschung ( <a href="http://www.toolpool-gesundheitsforschung.de">www.toolpool-gesundheitsforschung.de</a> )
TOS	Taled Open Studio; ETL-Software ( <a href="http://www.talend.com">www.talend.com</a> )
TransCelerate	TransCelerate Biopharma Inc.; gemeinnütziger Zusammenschluss von Pharmaherstellern, um sich gegenseitig „pre-competitive“ zu unterstützen ( <a href="http://transceleratebiopharmainc.com">http://transceleratebiopharmainc.com</a> )
tranSMART	Webbasierte Knowledge-Management-Plattform für die wissenschaftliche Hypothesenentwicklung auf Basis von Beziehungen zwischen phänotypischen und genetischen Daten ( <a href="http://transmartfoundation.org">http://transmartfoundation.org</a> )
TS	Technical Specification (ISO)
TTP	Trusted Third Party
TVL	Tarifvertrag für den öffentlichen Dienst der Länder
UCT	Universitäres Centrum für Tumorerkrankungen Frankfurt ( <a href="http://www.uct-frankfurt.de">www.uct-frankfurt.de</a> )
UCUM	Unified Code for Units of Measure ( <a href="http://unitsofmeasure.org">http://unitsofmeasure.org</a> )
UICC	International Union against Cancer ( <a href="http://www.uicc.org">www.uicc.org</a> )
UK	United Kingdom
UK	Universitätsklinik(en)
ULD	Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein ( <a href="http://www.datenschutzzentrum.de">www.datenschutzzentrum.de</a> )
UMLS	Unified Medical Language System ( <a href="http://www.nlm.nih.gov/research/umls">www.nlm.nih.gov/research/umls</a> )
UMTS	Universal Mobile Telecommunications System, Mobilfunkstandard
UniProt	Universal Protein Resource; Sammlung von Informationsressourcen zu Protein-Sequenzen und Annotationsdaten ( <a href="http://www.uniprot.org">www.uniprot.org</a> )

URI	Uniform Resource Identifier
URL	Uniform Resource Locator
USB	Universal Serial Bus: Standardisiertes Kommunikationssystem zur Verbindung von Computern mit Zusatzgeräten
USHIK	United States Health Information Knowledgebase der AHRQ ( <a href="http://ushik.ahrq.gov">http://ushik.ahrq.gov</a> )
UTD	University of Texas at Dallas
VCF	Variant Call Format, Textdatei-Format für die Speicherung von Informationen zu Genompositionen und zugehörigen Metadaten
VISTA	Webbasierte Studiensoftware der EORTC
Vivli	Center for Global Clinical Research Data; Projekt des MRCT ( <a href="http://mrctcenter.org/projects/vivli">http://mrctcenter.org/projects/vivli</a> )
VPH	Virtual Physiological Human
VUD	Verband der Universitätsklinika Deutschlands ( <a href="http://www.uniklinika.de">www.uniklinika.de</a> )
VV	Verification and Validation activity
W3C	World Wide Web Consortium ( <a href="http://www.w3.org">www.w3.org</a> )
WES	Whole Exome Sequencing
WG	Working Group
WGS	Whole Genome Sequencing
WHO	World Health Organization ( <a href="http://www.who.org">www.who.org</a> )
WMA	The World Medical Association; Weltärztebund ( <a href="http://www.wma.net">www.wma.net</a> )
WP	Working Party der Europäischen Arbeitsgruppe zum Datenschutz gemäß Artikel 29 der Richtlinie 95/46/EG
WR	Wissenschaftsrat ( <a href="http://www.wissenschaftsrat.de">www.wissenschaftsrat.de</a> )
WWW	World Wide Web
XML	extensible Markup Language
XNAT	Extensible Neuroimaging Archive Toolkit: Open source imaging informatics platform der Neuroinformatics Research Group der Washington University ( <a href="http://xnat.org">http://xnat.org</a> )
YODA	Yale University Open Data Access; Projekt und Portal zur Bereitstellung von Daten aus klinischen Forschungsvorhaben ( <a href="http://yoda.yale.edu">http://yoda.yale.edu</a> )
ZBR	Zentrales Biobprobenregister (im DZIF)
ZeBanC	Zentrale Biomaterialbank der Charité; Förderprojekt des BMBF im Rahmen der Nationalen Biobanken-Initiative ( <a href="http://biobank.charite.de">http://biobank.charite.de</a> )
Zenodo	Open Science Plattform für Veröffentlichungen, Forschungsergebnisse und Forschungsdaten; betrieben vom CERN und unterstützt von der EU ( <a href="https://zenodo.org">https://zenodo.org</a> )
ZIP	Standardisiertes Format zur Komprimierung von Dateien
ZKS	Zentrum für Klinische Studien





Patientenorientierte medizinische Forschung ist ohne die Unterstützung durch eine geeignete IT-Infrastruktur heute nicht mehr denkbar. Seit über 15 Jahren begleitet die TMF den digitalen Wandel in der medizinischen Forschung, der mit der Digitalisierung der Gesundheitsversorgung in engem Zusammenhang steht. Mit den Berichten des IT-Reviewing-Boards veröffentlicht die TMF seit 2013 eine kontinuierliche Bestands- und Bedarfsanalyse als Grundlage für die strategische Weiterentwicklung der IT-Infrastrukturen in der medizinischen Forschung.

In den ersten beiden Berichten sind die verschiedenen Datenquellen in den Blick genommen worden: von klinischen Studien über Register und Kohorten bis zur Omics-Technologie und der sekundären Nutzung klinischer Behandlungsdaten. Für jeden dieser Bereiche und die Querschnittsfunktion des ID-Managements wurde dargestellt, welche Lösungen möglich sind bzw. schon eingesetzt werden. Wie eine IT-Architektur für die Integration dieser heterogenen Daten gestaltet sein muss und welche Softwarelösungen hierfür schon vorhanden sind, ist als Schwerpunktthema im dritten Bericht aufgegriffen worden. Einzelne Aspekte wie Metadatenmanagement, Anonymisierung von Forschungsdaten und Cloud-Computing wurden in spezifischen Kapiteln vertieft behandelt. In der jetzt vorliegenden vierten Ausgabe wurde die bisherige Struktur und Kapitelteilung beibehalten. Alle Kapitel wurden umfassend aktualisiert und vermehrt Querbezüge der Kapitel untereinander hergestellt.

ISSN 2363-5754

ISBN 978-3-89838-723-1



9 783898 387231



[www.aka-verlag.com](http://www.aka-verlag.com)

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

