



# IT-Infrastrukturen in der patientenorientierten Forschung

Aktueller Stand und Handlungsbedarf

2014

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

2014

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 – 2014**

## **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.

© 2014, 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

Satz: Anja Brysch, Caroline Kryzecki

Verlag: Akademische Verlagsgesellschaft AKA GmbH,  
Postfach 41 07 05, 12117 Berlin | [www.aka-verlag.com](http://www.aka-verlag.com)

Druck: Bookstation GmbH, Anzing

Printed in Germany

ISSN: 2363-5754

ISBN: 978-3-89838-701-9

# 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, Daniela Skrowny  
*Institut für Medizinische Informatik, Universitätsmedizin Göttingen*  
sowie Prof. Dr. Tim Reißbarth  
*Institut für Medizinische Statistik, Universitätsmedizin Göttingen*

**Prof. Dr. Frank Ückert, Prof. Dr. Klaus Pommerening**

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

**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 Andreas Beck  
*Lehrstuhl für Medizinische Informatik, Medizinische Fakultät der  
Friedrich-Alexander-Universität Erlangen-Nürnberg*

**Ronald Speer**

*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**

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

<b>Einleitung</b>	<b>9</b>
-------------------	----------

---

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

---

<b>Identitätsmanagement</b>	<b>19</b>
-----------------------------	-----------

---

Anwendungsbereich	19
Betreiber, Anwender, Nutzer	20
Rahmenbedingungen	20
Kategorisierung der IT-Landschaft	22
Aktueller Bestand	23
Bewertung und Handlungsbedarf	24
Literatur	27

<b>Klinische Studien</b>	<b>29</b>
--------------------------	-----------

---

Anwendungsbereich	29
Betreiber, Anwender, Nutzer	30
Rahmenbedingungen	31
Kategorisierung der IT-Landschaft	32
Aktueller Bestand CDMS	34
Aktueller Bestand an Systemen für die Verwaltung von Studienteilnehmern	34
Bewertung und Handlungsbedarf	35
Literatur	37

<b>Register, Kohorten und Data Repositories</b>	<b>41</b>
---	-----------

---

Anwendungsbereich	41
Rahmenbedingungen	42
Kategorisierung der IT-Landschaft	43
Anforderungen an Kohortenstudien & Register IT (KoRegIT)	44
Aktueller Bestand	47
Bewertung und Handlungsbedarf	53
Danksagung	55
Literatur	56

## **Erhebung, Management und Verarbeitung digitaler Bilder** **59**

---

Anwendungsbereich	59
Betreiber, Anwender, Nutzer	61
Rahmenbedingungen	62
Kategorisierung der IT-Landschaft	63
Aktueller Bestand und Erhebungsergebnisse	64
Bewertung und Handlungsbedarf	66
Danksagung	68
Weiterführende Literatur	68

## **Biobanken** **69**

---

Anwendungsbereich	69
Betreiber, Anwender, Nutzer	70
Rahmenbedingungen	71
Kategorisierung der IT-Landschaft	74
Aktueller Stand	75
Bewertung und Handlungsbedarf	80
Literatur	82

## **Molekularbiologische Daten aus Hochdurchsatz-Analysen** **85**

---

Anwendungsbereich	85
Betreiber, Anwender, Nutzer	87
Rahmenbedingungen	88
Kategorisierung der IT-Landschaft	95
Aktueller Bestand	96
Bewertung und Handlungsbedarf	98
Danksagung	100
Literatur	101

## **Mobile IT-Werkzeuge** **105**

---

Anwendungsbereich und Eingrenzung	105
Betreiber, Anwender, Nutzer	107
Rahmenbedingungen	108
Kategorisierung der IT-Landschaft	110
Aktueller Stand	111
Bewertung und Handlungsbedarf	115
Literatur	116

## **Sekundärnutzung klinischer Daten** **119**

---

Anwendungsbereich	119
Betreiber, Anwender, Nutzer	121
Rahmenbedingungen	122
Kategorisierung der IT-Landschaft	129
Aktueller Stand	133
Bewertung und Handlungsbedarf	136
Literatur	139

## **Abkürzungsverzeichnis** **145**

---

# Einleitung

Wer IT-Infrastrukturen in der Forschung auf- oder ausbauen möchte, steht vor einer großen Zahl von Fragen und Entscheidungen, z. B.: Wie sieht der zu unterstützende Anwendungsfall genau aus? Welche Qualitätsparameter müssen beispielsweise von einer Biomaterialverwaltungssoftware erfasst und verarbeitet werden können? Welche gesetzlichen Rahmenbedingungen sind zu beachten? Dürfen die aus einer Probe gewonnenen Sequenzdaten mit demselben Ordnungskriterium oder derselben ID gespeichert werden wie die zugehörigen klinischen Daten? Welche Softwaresysteme müssen nach den Kriterien der Good Clinical Practice (GCP) validiert werden? Wie sollen die zu erfassenden Daten kodiert werden? Kann die von einem Laborsystem vorgegebene Kodierung übernommen werden, oder sollte besser auf einen internationalen Standard gemappt werden? Welche Schnittstellen müssen von den Softwaresystemen unterstützt werden? Muss meine Biomaterialverwaltungssoftware Lager- oder Aliquotierroboter ansteuern können? Welche Datenmengen müssen wie lange gespeichert werden können? Wieviel Speicherplatz wird pro Proband für eine Ganzgenomsequenzierung benötigt? Mit welcher Bandbreite müssen welche Systeme angebunden werden?

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: Welche Finanzmittel werden für die IT in den nächsten Jahren oder in dem zu beantragenden Projekt benötigt? Wieviel Entwicklungskapazität wird am Standort für Individuallösungen benötigt? Wie müssen die Forschungsfragestellungen an die gesetzlichen Rahmenbedingungen angepasst werden? Wie kann die IT das unterstützen? Muss ich eine eigene Biobank aufbauen, oder kann ich mich einer bestehenden anschließen? Kann eine 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? Welche IT-Infrastrukturen und welche Standards können vorausgesetzt 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. 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 Pflege dieser Bestandsaufnahme von IT-Strukturen wurde mit Unterstützung des Bundesministeriums für Bildung und Forschung (BMBF) ein IT-Reviewing-Board gegründet, welches diese Sichtung vornimmt und in Form eines öffentlichen Berichts zur Verfügung stellt. Die zweite Ausgabe des Reports zu „IT-Infrastrukturen in der patientenorientierten Forschung in Deutschland“ liegt mit dieser Publikation vor. Das Board setzt sich aus Experten für IT-Infrastrukturen in der medizinischen Forschung und Versorgung zusammen und 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. Alle Mitglieder des IT-Reviewing-Boards haben in der einen oder anderen Form, meist in zentralen Rollen, über viele Jahre in den Arbeitsgruppen, Projekten und anderen Querschnittsaktivitäten der TMF mitgewirkt.

Der Report des IT-Reviewing-Boards wird jährlich mit wechselnden Schwerpunkten, Themen und Autoren fortgeschrieben. Schwerpunkt der ersten Ausgabe war die Datenerhebung (s. Abb. 1).

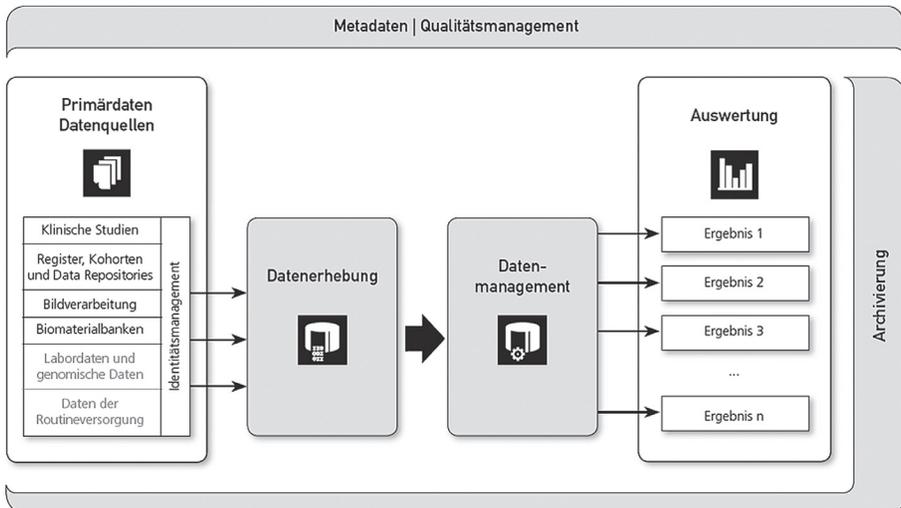


Abbildung 1: Generische Darstellung der Datenverarbeitung in der patientenorientierten Forschung von den Datenquellen bis zur Auswertung und Archivierung. Auf die Verarbeitung von molekulargenetischen und sekundär genutzten Daten aus der Routineversorgung wird in diesem Bericht erstmals eingegangen.

Die aktuelle zweite Ausgabe schreibt diesen Schwerpunkt fort, enthält jedoch erstmals auch jeweils ein Kapitel zur Erhebung und Verarbeitung molekularbiologischer Daten sowie zur Sekundärnutzung klinischer Daten für die Forschung. Alle weiteren Kapitel wurden umfassend aktualisiert.

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, der aktuelle Bestand und schließlich eine Bewertung mit dem sich daraus eventuell ergebenden Handlungsbedarf.

## 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“, „Medizintechnik“ 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.

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.



# Zusammenfassung

Patientenorientierte medizinische Forschung ist ohne die Unterstützung durch eine geeignete IT-Infrastruktur heute nicht mehr denkbar. Seit 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.

Der Bericht 2012/2013 hatte – mit Blick auf die Primärdatenquellen – 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 vorliegende zweite IT-Report nimmt mit der Sekundärdatennutzung und den molekularbiologischen Daten aus Hochdurchsatz-Analysen (Omics-Technologien) zwei zusätzliche Datenquellen in den Blick. Dabei wird auch auf die wichtige Verzahnung von Forschung und Versorgung eingegangen. Es kristallisieren sich über die verschiedenen betrachteten Forschungsfelder hinweg die folgenden Anforderungen und Empfehlungen heraus, die für eine Integration und Weiterentwicklung der Forschungs-IT in der Medizin kritisch sind:

- ▶ **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. Eine Reihe von Plattformen, Registern und Projekten ist hierzu in den vergangenen Jahren bereits auf den Weg gebracht worden (z.B. Deutsches Biobanken-Register, Register für Register, Nationales Metadaten-Repository, „IT-Service-Portal“). Allerdings besteht nach wie vor die Schwierigkeit, diese Infrastrukturen aus der Projektförderung in den langfristigen Betrieb zu überführen. Hier sind geeignete Finanzierungs- und Betriebsmodelle sowie Anreizsysteme zu entwickeln. Dies kann nur mit Unterstützung der Politik und der Förderorganisationen geschehen.
- ▶ **Standardisierung:** Die Nutzung von Standards für die Erfassung, Verarbeitung und Weitergabe von Daten ist unerlässlich. Das gilt für Daten in klinischen Studien und Registern, für Daten aus molekularen Analysen, für Bilddaten sowie auch für Daten und Proben in Biobanken. Auch für die Nutzung von Daten aus der Versorgung für die Beantwortung von Forschungsfragen ist eine standardisierte Dokumentation eine wesentliche Voraussetzung. Die Autoren des Berichts fordern, dass bereits existierende Standards noch viel stärker als bisher bekannt gemacht, geschult und genutzt werden.

- ▶ **Interoperabilität:** Medizinischer Fortschritt beruht immer stärker auch auf der Möglichkeit, Informationen aus verschiedenen Sektoren und IT-Systemen zusammenzuführen. Das scheitert allerdings oftmals noch an der fehlenden Interoperabilität der Systeme und an einem Mangel an geeigneten Schnittstellen. Neben der Harmonisierung der Daten- und Übertragungsstandards ist insbesondere auch eine semantische Annotation der Daten der beteiligten Systeme notwendig. Perspektivisch wird hierfür der Einstieg in die systematische Evaluation und Nutzung einer Referenz-Terminologie wie SNOMED-CT empfohlen. Eine solch tiefgreifende und die Einbindung vieler Parteien aus Forschung und Versorgung erfordernde Maßnahme kann nicht ohne den Aufbau einer geeigneten nationalen Begleit- und Steuerstruktur 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.
- ▶ **Aufbau von Dienstleistungsangeboten:** Vermehrt wird der Bedarf geäußert, im Bereich der akademischen medizinischen Forschung zentral angebotene Dienstleistungen zu moderaten Preisen zu etablieren. Dies gilt sowohl für die Nutzung von Softwarekomponenten als auch für das Identitätsmanagement im Rahmen der benötigten Datenschutzkonzepte oder für Analyse-Leistungen im Bereich der Omics-Technologien. 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 Hochdurchsatz-Genotypisierungsdaten (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.

## Neue Themenschwerpunkte 2014/2015

### **Sekundärdatennutzung**

Mit zahlreichen relevanten Anwendungsfällen gewinnt das Thema der Nutzung von Daten aus der Versorgung für die Forschung zunehmend an Bedeutung. Jedoch besteht, wie ein von der TMF eingeholtes Rechtsgutachten zur Sekundärdatennutzung zeigt, derzeit in Deutschland eine uneinheitliche und in Teilen unklare Rechtslage. Ein einheitlicher und adäquater Rechtsrahmen, der allen Akteuren ausreichende Rechtssicherheit bietet, müsste dringend geschaffen werden. Ob die derzeit in Diskussion befindliche europäische Datenschutz-Grundverordnung zur Verbesserung der Situation beiträgt, ist noch offen. Große Hürden bestehen zudem in der fehlenden Transparenz und einer ungenügenden Standardisierung der Versorgungsdokumentation. So ist es kaum möglich zu beurteilen, welche Datenelemente in verschiedenen Datenquellen verfügbar sind und daher für eine gemeinsame wissenschaftliche Auswertung überhaupt in Frage kommen. Von wissenschaftlicher Seite wird deshalb ein freier Zugang zu den Metadaten im Gesundheitswesen gefordert. Die erforderlichen Abstimmungs- und Standardisierungsprozesse sind durch eine neutrale Stelle auf nationaler Ebene zu begleiten und zu moderieren.

### **Omics-Technologien**

Neu ist ebenfalls ein Kapitel zu IT-Infrastrukturen der so genannten Omics-Methoden. Während im Biobanking bereits professionelle Lösungen für die Dokumentation von Workflows sowie für das Proben- und Datenmanagement existieren, besteht weiterhin großer Bedarf an der Entwicklung und Einführung von adäquaten, standardisierten IT-Lösungen für Omics-Labore. Der Datenaustausch zwischen den Akteuren erfolgt bisher in den wenigsten Fällen automatisiert. Hier besteht großer Handlungsbedarf, um die nachhaltige Verfügbarkeit der Daten gemäß regulatorischer Vorgaben und eine Reproduzierbarkeit von Ergebnissen zu erreichen. Bei weiter sinkenden Kosten für molekulare Hochdurchsatz-Analysen werden künftig massive Probleme bei der langfristigen Speicherung der Daten auftreten. Dringend müssen Verfahren und Regelungen

etabliert werden, die klären, wer die Daten nach den Regeln der guten wissenschaftlichen Praxis aufbewahren kann und muss. Der Umgang mit und die Analyse von Omics-Daten muss in der Ausbildung von Forschern und Ärzten stärker verankert werden. Eine enge Abstimmung zwischen den verschiedenen Omics-Disziplinen und der Bioinformatik ist notwendig. Für die Integration von NGS-Analysen in die klinische Routine müssen konkrete Konzepte entwickelt und umgesetzt werden.

### Fortschritte seit 2012/2013

Die Bestandsaufnahme und die Empfehlungen zu den Themen Identitätsmanagement, klinische Studien, Register, Bildverarbeitung, Biobanken und mobile IT sind für den vorliegenden Bericht aktualisiert worden. Neben den oben zusammengefassten Bedarfen und Forderungen beschreiben die Autoren die Fortschritte, die sich seit dem IT-Report 2012/2013 ergeben haben:

- ▶ Die neuen modularen und an die aktuellen Anforderungen angepassten generischen Datenschutzkonzepte der TMF sind 2014 in Buchform erschienen. Ihre Nutzung wird seither in verschiedenen Schulungsveranstaltungen vermittelt.
- ▶ Die TMF baut derzeit ein Web-Portal auf, das Informationen zum Aufbau und Betrieb von IT-Komponenten für die medizinische Forschung zusammenführen und aufbereiten wird. Dieses Angebot wird in enger Verzahnung mit der Arbeit des IT-Reviewing-Boards konzipiert.
- ▶ Der Kerndatensatz für ein Register der Register ist abgestimmt und verabschiedet worden. Notwendig ist jetzt eine tragfähige Finanzierung für ein solches Übersichtsregister.
- ▶ Es liegt ein Anforderungskatalog zur IT-Unterstützung von Kohorten und Registern vor, der für die Planung entsprechender Infrastrukturen genutzt werden kann. Zudem ist 2014 auch eine Neuauflage der TMF-Leitlinie zur Datenqualität in Kohortenstudien und Registern veröffentlicht worden.
- ▶ Die TMF hat 2014 eine Kooperationsvereinbarung mit IHE-Deutschland getroffen, bei der es darum geht, im Sinne der Verzahnung von medizinischer Forschung und Versorgung die Nutzung von IHE-Profilen und Konzepten in der Forschung zu prüfen und voranzutreiben.
- ▶ Im Biobanking etablieren sich Nutzer-Gruppen (user groups) zu den im akademischen Umfeld überwiegend genutzten Softwaresystemen. Die Ergebnisse aus dem Austausch der verschiedenen Gruppen sollten weiterhin in der Arbeitsgruppe Biomaterialbanken der TMF zusammengeführt werden.

- ▶ Die mobile Datenerfassung in der Forschung erfährt durch die Verbreitung leistungsfähiger Smartphones einen enormen Schub. Die standardisierten Infrastrukturen nur weniger Systemanbieter vereinfachen die Entwicklung von Apps, die von vielen Patienten und Probanden zur eigenständigen Datenerfassung genutzt werden können. Gleichzeitig bringt die Nutzung der relevanten Systemumgebungen erhebliche Datenschutzrisiken mit sich. Hier sind dringend pragmatische Lösungsansätze und Handlungsempfehlungen zu erarbeiten.



# Identitätsmanagement

*Klaus Pommerening, Frank Ückert*

## Anwendungsbereich

Gegenstand dieses Abschnitts ist das Identitätsmanagement für Patienten oder Probanden in der medizinischen Forschung. Dieses ist ein Kernthema des Datenschutzes und überall dort relevant, wo mit patientenbezogenen Daten geforscht wird. Es dient dazu, die Identität der Betroffenen durch die Verwendung von Pseudonymen zu verbergen und 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 hier 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 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 aus der rechtlichen Grauzone geführt, 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 Datenschutz-

beauftragten 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. Die Anwendung geschieht in der Regel implizit, etwa weil das Identitätsmanagement in Software zum Electronic Data Capturing (EDC) eingebaut ist, z. B. mit Hilfe der von der gemeinsam von der Universität Münster und der TMF zur Verfügung gestellten Datenschutz-Bibliothek<sup>1</sup> zum Klinischen Modul des TMF-Datenschutzleitfadens [1] (vergl. Modell A der ersten generischen Datenschutzkonzepte der TMF [2]). Alternativ können Funktionen des ID-Managements als Netzdienst in die Systemkommunikation eingebunden sein, wie es z. B. auch der PID-Generator<sup>2</sup> und der Pseudonymisierungsdienst<sup>3</sup> der TMF sowie als Weiterentwicklung die Mainzliste<sup>4</sup> unterstützen. Direkte Anwender sind daher im Weiteren nur Verbundpartner mit Treuhänderfunktion, wobei die direkte Anwendung bei diesen sich im Regelfall auf Depseudonymisierungsprozesse beschränkt.

Implizit sind aber 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 die Dienste fast durchweg in Eigenregie der Forschungsverbände betrieben und die vorhandenen Konzepte oft eher zur Orientierung für eigene Implementierungen genutzt.

## Rahmenbedingungen

Die wesentlichen **rechtlichen Rahmenbedingungen** für das Identitätsmanagement liegen im Datenschutzrecht und im Arztrecht (das zusammen mit dem Strafrecht die ärztliche Schweigepflicht fest schreibt). Das Datenschutzrecht ermöglicht im Regel-

---

1 <http://www.tmf-ev.de/Produkte/P014012>

2 <http://www.tmf-ev.de/Produkte/P015011>

3 <http://www.tmf-ev.de/Produkte/P000011>

4 <http://www.unimedizin-mainz.de/imbei/informatik/opensource/mainzliste.html>

fall 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. Auch der derzeit vorliegende Entwurf für eine neue EU-Datenschutzverordnung [4] gibt den Weg eines pseudonymisierten Identitätsmanagements vor.

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>5</sup> gegeben. Hier liegen generische Musterlösungen für medizinische Forschungsverbünde vor, die schon in vielen Fällen umgesetzt wurden und zu einem akzeptierten und mit Datenschützern abgestimmten Datenschutzkonzept des jeweiligen Forschungsverbunds führten.

Für den verteilten Betrieb der in den Konzepten vorgegebenen Dienste des ID-Managements werden geeignete unabhängige Dienstleister benötigt, z. B. Medizininformatische (MI) 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 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 zugehöriger IDs fehlen oder für eher versorgungsorientierte Anwendungsfälle konzipiert wurden, wie z. B. im Rahmen von HL7-Nachrichten oder in den IHE-Profilen PIX/PDQ.

---

5 <https://www.datenschutz.hessen.de/dg011.htm#entry4196>

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. Den Verantwortlichen für die Verbände muss vermittelt werden, dass erst ein sorgfältig geplantes und sicher implementiertes Identitätsmanagement eine datenschutzgerechte Zusammenarbeit im Verbund über institutionelle Grenzen hinweg ermöglicht und dass es nicht eine zusätzliche bürokratische Hürde darstellt, sondern dabei hilft, 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 Infrastruktur ebene 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.

Fazit: Damit das Identitätsmanagement in heterogene IT-Landschaften integriert werden kann, sollte die Implementation auf zentralen Diensten beruhen, die als Web-Services über standardisierte Schnittstellen genutzt werden können.

## Aktueller Bestand

Zurzeit werden in den 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, für das Forschungsmodul nur im Ansatz.
- ▶ 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 zum Datenschutzleitfaden, das eine Überarbeitung der bisherigen generischen Datenschutzkonzepte zur Aufgabe hatte, 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>6</sup>
- ▶ Das für das Register des Netzwerks für Ichthyosen und verwandte Verhornungsstörungen (NIRK) in Mainz entwickelte Werkzeug „Mainzliste“, die im Mai 2013 unter einer Open-Source-Lizenz (AGPLv3) von der Universitätsmedizin Mainz verfügbar gemacht wurde,<sup>7</sup> kann als geeignete Weiterentwicklung des PID-Generators angesehen werden. Es wird inzwischen auch in mindestens 13 Verbänden eingesetzt. Die Rückmeldungen, soweit existent, sind durchweg positiv.

---

6 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.

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

- ▶ Ein ähnliches Leistungsspektrum bieten die in Greifswald im Rahmen des MOSAIC-Projekts veröffentlichten Werkzeuge E-PIX (zum Identitätsmanagement) und gPAS (zur Pseudonymverwaltung), die ebenfalls unter einer Open-Source-Lizenz zur Verfügung gestellt werden<sup>8</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>9</sup>:

- ▶ PID-Generator mit Patientenliste, fehlertolerantem Record-Linkage, deutscher Phonetik nach Michael [5] und Postel [6] sowie SOAP-Schnittstelle,
- ▶ Pseudonymisierungsdienst für die zweite Stufe einer doppelten Pseudonymisierung mit je einer Schnittstelle zu Studiendatenbanken und Forschungsdatenbanken,
- ▶ Interoperabilität mit EDC-Software – etwa per proprietärem PID-Dispatcher in SecuTrial 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 neuen Datenschutzleitfadens der TMF) für javascriptfähige Browser. SecuTrial wird in dieser Form z. B. in den Kompetenznetzen „Parkinson“ und „Angeborene Herzfehler“ genutzt. Für den Einsatz der Datenschutzbibliothek „Modell A“ kann das Register des Netzwerks für Ichthyosen und verwandte Verhornungsstörungen (NIRK) als Referenz dienen.

Die in Mainz (Mainzliste) und Greifswald (E-PIX und gPAS) verfügbaren Werkzeuge werden auch als webbasierte Dienstleitungen angeboten.

## 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 angebotenen Werkzeuge kann man resümieren:

---

8 <https://mosaic-greifswald.de/werkzeuge-und-vorlagen.html>

9 siehe [www.tmf-ev.de/produkte](http://www.tmf-ev.de/produkte)

- ▶ Der PID-Generator wird in vielen Verbänden eingesetzt und hat seine Nützlichkeit und Stabilität bewiesen. Für neu beginnende Forschungsverbände sind die Mainzelliste oder E-PIX/gPAS als Weiterentwicklungen geeignet und haben sich bereits bewährt.
- ▶ Der Pseudonymisierungsdienst wird derzeit noch nicht produktiv eingesetzt.
- ▶ 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, besteht folgender Handlungsbedarf:

- ▶ Der Pseudonymisierungsdienst muss als generisch einsetzbares Werkzeug zur Umschlüsselung von Pseudonymen und als Grundbaustein einer informationellen Gewaltenteilung überarbeitet oder neu entwickelt werden. Um eine robuste und praktikabel nutzbare Software zu erhalten, müssen spätere Betreiber eng in die Entwicklung eingebunden werden.
- ▶ Die Mainzelliste als Nachfolgeprodukt des PID-Generators oder vergleichbar mächtige Softwarekomponenten (E-PIX, gPAS) müssen dahingehend aktualisiert und erweitert werden, dass alle für eine Umsetzung des Datenschutzleitfadens der TMF notwendigen Funktionen unterstützt werden und dass eine solche Softwarekomponente in aktuellen Systemumgebungen sicher eingesetzt werden kann.
- ▶ Weiterentwicklung der Konzepte und Umsetzungen, insbesondere
  - ▶ Verbesserung anhand der Rückmeldungen von Anwenderworkshops
  - ▶ Internationalisierung
  - ▶ Einbindung oder Anbindung eines Kontaktmanagements (CRM)
  - ▶ Einbindung oder Anbindung eines Einwilligungsmanagements
  - ▶ Interoperabilität zwischen den exemplarischen Werkzeugen in den Bereichen Pseudonymisierung, Kontaktmanagement, Einwilligungsmanagement und Web-basierte Registerwerkzeuge
  - ▶ Entwicklung sicherer Anonymisierungsfunktionen (z. B. zur  $k$ -Anonymisierung)<sup>10</sup>
- ▶ Aufbau einer Interoperabilitätsplattform für Infrastrukturkomponenten und Integration der Werkzeuge zum Identitätsmanagement in diese. Die Förderung eines entsprechenden Projektes (TAILOR) wird z. Z. beantragt.
- ▶ 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.

---

<sup>10</sup> Ein Softwareframework zur Umsetzung einer  $k$ -Anonymisierung wird als TMF-Produkt bereits zur Verfügung gestellt, s. [www.tmf-ev.de/Produkte/P100201](http://www.tmf-ev.de/Produkte/P100201)

- ▶ 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 Forschungsverbände 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 Forschungsverbände gegenseitig Dienstleistungen anbieten. Es ist bereits absehbar, dass sich aus dieser Problematik heraus ein dringender Standardisierungsbedarf ergibt, 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 Forschungsverbände nachhaltig gefördert.
- ▶ Ebenfalls wünschenswert ist eine systematische Evaluation der bereits eingesetzten Werkzeuge (Mainzliste, E-PIX, gPAS).
- ▶ Langfristig sollten internationale Standards für die Nutzung von ID-Management-Funktionen entwickelt werden, die eine standardisierte Anbindung eines ID-Managements an proprietäre Softwaresysteme in der Versorgung und der Forschung ermöglichen, z. B. auf der Basis von ISO/TS 25237<sup>11</sup> oder des IHE Healthcare Data Pseudonymization Handbook<sup>12</sup>.

---

11 derzeit im Review

12 derzeit als Internal Draft, s. [ftp://ftp.ihe.net/IT\\_Infrastructure/iheityr10-2012-2013/Technical\\_cmte/WorkItems/PseudonymizationWhitePaper/](ftp://ftp.ihe.net/IT_Infrastructure/iheityr10-2012-2013/Technical_cmte/WorkItems/PseudonymizationWhitePaper/)

## Literatur

1. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten – Generische Lösungen der TMF 2.0*. 2014, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
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, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
3. Becker, R., Ihle, P., Pommerening, K., Harnischmacher, U. *Ein generisches Datenschutzkonzept für Biomaterialbanken (Version 1.0)*. 2006, <http://www.tmf-ev.de/produkte/P010021> (Abruf: 2014-06-10).
4. EC *Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutz-Grundverordnung)*. 2012. Europäische Kommission, [http://ec.europa.eu/justice/data-protection/document/review2012/com\\_2012\\_11\\_de.pdf](http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_de.pdf) (Abruf: 2014-06-10).
5. Michael, J., *Doppelgänger gesucht. Ein Programm für kontextsensitive phonetische Textumwandlung*. c't, 1999. **25**: S. 252–261.
6. Postel, H.J., *Die Kölner Phonetik. Ein Verfahren zur Identifizierung von Personennamen auf der Grundlage der Gestaltanalyse*. IBM-Nachrichten, 1969. **19**: S. 925–931.



# Klinische Studien

*Christian Ohmann, Ronald Speer*

## 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. Die 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). Gemäß der Schwerpunktsetzung dieses Berichts im Bereich der Datenerhebung werden im Folgenden 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 hohe Anforderungen an die Softwarelösungen stellen.

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 akademische 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.

## Betreiber, Anwender, Nutzer

Der Einsatz von Studiensoftware kann in Studienzentren auf der Basis eigener Serverinstallationen oder auch im Rahmen eines Application Service Providing 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 verschiedenen 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, bedingt durch die Aussicht auf eine als Service verfügbare preiswertere und flexiblere IT-Infrastruktur und eine Reduzierung der Anforderungen an Inhouse-Systeme und Personal, erheblich gestiegen. Konsequenterweise gibt es immer mehr CDMS-Anbieter, die ihre Systeme als SaaS (Software as a Service) und über cloud-basierte Datenzentren anbieten (z. B. OpenClinica, Medidata rave). 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].

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 etc.) zu sehen. Eine Bestandsaufnahme der TMF im Rahmen eines Projektes zeigte, dass hier sehr heterogene Systeme zum Einsatz kommen.<sup>13</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 folgenden Rollen unterscheiden: Sponsoren, Prüfarzte und Mitglieder der Prüfgruppe sowie Personal der Studienzentren (z. B. Projekt- und Datenmanager, Monitore). Bei Studien nach Arzneimittelgesetz (AMG) [2] oder Medizinproduktegesetz (MPG) [3] kommt dem Sponsor eine zentrale Rolle zu,

---

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

da er für die Planung, Durchführung und Auswertung einer klinischen Studie die Gesamtverantwortung trägt. Prüfärzte 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 Grundlage der klinischen Studienforschung ist in Deutschland das AMG für Studien mit Arzneimitteln und das MPG für Studien mit Medizinprodukten [2, 3]. 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) [4]. Dabei steht 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 (z. B. [5]). 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. 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 [6, 7]. 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 Pilotprojektes erfolgreich implementiert werden konnte. Mittlerweile wurde mit ECRIN-ERIC eine europäische Infrastruktur mit legalem Status aufgebaut. ECRIN-ERIC hat die Zertifizierung von Datenzentren in den Arbeitsplan aufgenommen und einen ersten Call veröffentlicht. Derzeit befinden sich 5 Zentren aus 5 EU-Ländern im Zertifizierungsprozess. Der nächste Call erfolgt voraussichtlich im zweiten Quartal 2015.

Eine Arbeitsgruppe des KKS-Netzwerks hat unter Federführung des ZKS Köln aktuell Requirements für CDMS erarbeitet. Mit insgesamt 331 Anforderungen deckt der

Anforderungskatalog sowohl web-basierte als auch papier-basierte Studien ab. Neben den Kernanforderungen (eCRF-Implementierung, Anwender- und Rollenverwaltung, Datenerfassung, prüfung und korrektur, Berichte, Datenexport, Audit Trial) wurden auch Anforderungen für Zusatzfunktionen (Codierung von unerwünschten Ereignissen und Medikamenten, Berichtsgenerator, Planung und Vorbereitung von Monitoring-Visiten, Randomisation, Medikationsmanagement, SAE-Management, Studienmanagement) formuliert.

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 dieses Evaluierungsprojektes und mit BMBF-Mitteln konnten zwei Softwarelösungen in KKS/ZKS erfolgreich implementiert werden (MACRO, eResearchNetwork). 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) [8-10], aus Empfehlungen für die Validierung von CDMS-Systemen (z. B. [11]) sowie aus Standardisierungsbemühungen in diesem Feld (z. B. CDISC, [12]). 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 aktuell eine Spezifikation für einen Electronic Trial Master File (eTMF) entwickelt. Ziel dieses offenen und technologieneutralen Standards ist die Interoperabilität und der Datenaustausch von Trial Master Files bei klinischen Studien [13]. Nach wie vor ein Problem bei Standards für klinische Studien ist ihre mangelnde Akzeptanz und Anwendung.

## Kategorisierung der IT-Landschaft

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 Prüfzentren, 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 CDMS

Eine Umfrage zu den eingesetzten CDMS in 74 Zentren des ECRIN-Netzwerkes im Jahre 2010 zeigte folgende Verteilung in den Kategorien [14]:

- ▶ 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 [14].

In den KKS/ZKS werden derzeit vorwiegend MACRO und eClinical, aber auch secuTrial, Clincase, Marvin und Trialmaster eingesetzt.

Im Rahmen eines durchgeführten CDMS-Projekts der KKS-Arbeitsgruppe wurden Softwaresysteme von 19 Providern eruiert, z. B. Rave, TrialMaster, secuTrial, MACRO, OpenClinica, 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 [15]. Ähnliche Bemühungen, wenn auch in kleinerem Rahmen, gibt es z. B. in der Swiss Clinical Trial Organization mit der Einführung von SecuTrial [16].

## Aktueller Bestand an Systemen für die Verwaltung von Studienteilnehmern

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 Lösungen für das sogenannte Customer-Relationship-Management (CRM) sich nur unzureichend eignen und insbesondere die notwendigen zusätzlichen Funktionen im Rahmen klinischer Studien nur unzureichend unterstützen [17]. Auf der anderen Seite gibt es auch Probleme, vorhandene Patientenmanagementsysteme der beteiligten Prüfzentren zu nutzen. Neben den Funktionen für die datenschutzrechtlichen Anforderungen fehlen gerade hier einheitliche Schnittstellen.

## 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 eines laufenden, von der EU geförderten Projekts entwickelt die European Organisation for Research and Treatment of Cancer (EORTC) mit Beteiligung einzelner KKS ein CDMS für den akademischen Bereich (VISTA). VISTA ist eine Weiterentwicklung einer bestehenden und routinemäßig eingesetzten Software und wird nach Fertigstellung CDISC-, GCP- und 21 CFR Part 11-kompatibel ausgeliefert. Es wird erwartet, dass die erste Version von VISTA im Jahre 2015 zur Verfügung steht. Wünschenswert wären darüber hinaus einheitliche Supportangebote durch passende Dienstleister sowie günstige ASP-Angebote.

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.

Die immer komplexer werdenden Infrastrukturen mit einer 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.

Zudem gibt es beim Einsatz cloudbasierter 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; [18]). Dies Interesse ist sowohl in der Industrie als auch in der akademischen Forschung festzustellen und kann z. B. auch zusätzlich an die im Rahmen klinischer Studien erhobenen und ggf. weiter verwertbaren Bioproben gekoppelt sein. Nicht zuletzt diese Anforderungen haben auch zu den neuen modularen Datenschutzkonzepten der TMF geführt, die hierfür die notwendigen rechtlichen, technischen und organisatorischen Rahmenbedingungen beschreiben [19]. 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)
- ▶ Interoperabilität von CDMS mit EHR (laufende EU-Projekte: Transform, EHR4CR [20], p-medicine; abgeschlossenes BMBF-Projekt KISREK [21])
- ▶ Schnittstellen von CDMS zu Biobanken, Bilddaten, genetischen Daten, etc. (laufende 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.

---

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

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

- ▶ Entwicklung und Abstimmung von Anforderungen und Leitlinien zur Nutzung cloud-basierter Services für klinische Studien (z. B. CDMS als SaaS).
- ▶ 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*, submitted.
2. AMG, *Gesetz über den Verkehr mit Arzneimitteln (Arzneimittelgesetz – AMG) in der Fassung der Bekanntmachung vom 12. Dezember 2005 (BGBl. I S. 3394), das durch Artikel 1 der Verordnung vom 17. Dezember 2014 (BGBl. I S. 2222) geändert worden ist*, 1976/2014, [http://www.gesetze-im-internet.de/bundesrecht/amg\\_1976/gesamt.pdf](http://www.gesetze-im-internet.de/bundesrecht/amg_1976/gesamt.pdf) (Abruf: 2015-02-23).
3. MPG, *Medizinproduktegesetz in der Fassung der Bekanntmachung vom 7. August 2002 (BGBl. I S. 3146), das zuletzt durch Artikel 11 des Gesetzes vom 21. Juli 2014 (BGBl. I S. 1133) geändert worden ist*, 2005/2014, <http://www.gesetze-im-internet.de/bundesrecht/mpg/gesamt.pdf> (Abruf: 2015-02-23).
4. 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: 2015-02-23).
5. BDSG, *Bundesdatenschutzgesetz in der Fassung der Bekanntmachung vom 14. Januar 2003 (BGBl. I S. 66), das zuletzt durch Artikel 1 des Gesetzes vom 14. August 2009 (BGBl. I S. 2814) geändert worden ist*, 2003/2009, [http://www.gesetze-im-internet.de/bundesrecht/bdsg\\_1990/gesamt.pdf](http://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf) (Abruf: 2015-02-23).
6. 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.
7. 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.
8. 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: 2015-02-23).

9. 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: 2015-02-23).
10. 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: 2015-02-23).
11. TMF *V019-01 Systemvalidierung II. Systemvalidierung – Einsatz, Evaluierung und Erweiterung des Validierungspaketes*. 2005–2007, [http://www.tmf-ev.de/Themen/Projekte/V019\\_01\\_Systemvalidierung\\_II.aspx](http://www.tmf-ev.de/Themen/Projekte/V019_01_Systemvalidierung_II.aspx) (Abruf: 2015-02-23).
12. CDISC *Clinical Data Interchange Consortium*. 2014, <http://www.cdisc.org/> (Abruf: 2015-02-23).
13. OASIS *Electronic Trial Master File (eTMF) Specification Version 1.0, Committee Specification Draft 01 /Public Review Draft 01*. 2014. OASIS, <http://docs.oasis-open.org/etmf/etmf/v1.0/csprd01/etmf-v1.0-csprd01.html> (Abruf: 2015-02-23).
14. 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.
15. NCI *Implementation of the NCI's National Clinical Trials Network. NCI Perspective Article*. 2014, <http://www.cancer.gov/newscenter/newsfromnci/2014/NCTNlaunch>.
16. SCTO *Swiss Clinical Trial Organisation – Data Management. Promoting a nationally harmonised culture of high-quality clinical research*. 2015. Swiss Clinical Trial Organisation, [http://www.scto.ch/dms/SCTO/de/Publikationen/Factsheets/SCTO\\_PR\\_PUB\\_Factsheet\\_SCTO\\_DMWG\\_150121.pdf](http://www.scto.ch/dms/SCTO/de/Publikationen/Factsheets/SCTO_PR_PUB_Factsheet_SCTO_DMWG_150121.pdf) (Abruf: 2015-02-23).
17. 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.
18. 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.
19. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

20. Doods, J., Bache, R., McGilchrist, M., Daniel, C., Dugas, M., Fritz, F., Package, o.b.o.W., *Piloting the EHR4CR Feasibility Platform across Europe*. *Methods of Information in Medicine*, 2014. **53**(4): S. 264–268.
21. Schreiweis, B., Trinczek, B., Kopcke, 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, Kohorten und Data Repositories

*Jürgen Stausberg, Claudia Michalik*

## 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 Register auch einen Beitrag zur Gesundheitsversorgung, z. B. im Bereich der Qualitätssicherung und des Qualitätsmanagements. 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 bei ihren Mitgliedsverbänden analysiert. Die Beschreibung der gewonnenen Ergebnisse aus der Erhebung von 2012 stellt einen Schwerpunkt dieses Kapitels dar.

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“ [2, 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 eine Vielzahl von Zielen unterstützen, z. B. die Beschreibung epidemiologischer Zusammenhänge und Unterschiede, die Unterstützung von Qualitätssicherung und -verbesserung, die Unterstützung klinischer Forschung, die Evaluation und das Monitoring der Patientensicherheit, die Evaluierung der Wirksamkeit in der Versorgungsroutine, die ökonomische Evaluation, die Mindestmengenforschung sowie die Unterstützung der Versorgungsplanung. Die US-amerikanische Agency for Healthcare Research and Quality 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“ [3]. Das Einschlusskriterium bei Registern entspricht dem Outcome bei 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 [4]. 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.

## Rahmenbedingungen

Für Register, Kohorten und Data Repositories existieren nur wenige spezifische gesetzliche Regelungen. Umfassend geregelt ist der Bereich epidemiologischer Krebsregister durch Gesetze der Bundesländer. Durch das Gesetz zur Einführung klinischer Krebsregister ist auch für diesen Bereich eine gesetzliche Grundlage entstanden. Weitere Register-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 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, Kohorten und Data Repositories 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, Kohorten und Data Repositories [5]. Derzeit wandeln sich Register, Kohorten und Data Repositories von abgeschlossenen, national begrenzten Projekten zu internationalen, verteilten Plattformen. Hierdurch entstehen neue Herausforderungen, von der IT-Infrastruktur bis zur Rechtssicherheit.

## Kategorisierung der IT-Landschaft

Ein Markt für rechnergestützte Anwendungssysteme für Register, Kohorten und Data Repositories existiert (noch) nicht. Im Wesentlichen sind es vier Kernprozesse, für die der Einsatz spezifischer Anwendungssysteme in Frage kommt:

1. die Kommunikation mit den Studienzentren,
2. die Verwaltung der Probanden,
3. das Management der Daten von der Erhebung bis zum Monitoring sowie
4. die Auswertung und Nutzung der Daten.

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 [6]. Für die insbesondere bei Kohorten wichtige Verwaltung der Probanden werden bislang jedoch vorwiegend Eigenentwicklungen genutzt. Für das Management von Daten dominieren noch Studienmanagementsysteme bzw. EDC-Systeme 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 noch einfache 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, 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 [7]. Es bleibt abzuwarten, ob hieraus professionelle Lösungen entstehen. Abschließend bleibt zu erwähnen, dass für den Bereich klinischer Krebsregister einzelne Lösungen zumindest Teilaspekte aller vier oben genannten Kernprozesse abdecken. 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.

## 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 [8]. Dieser soll es ermöglichen, die Systemlandschaft von Kohorten und Registern in ihrem Bestand und hinsichtlich des Bedarfs besser zu beschreiben bzw. 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. Der Anforderungskatalog ist über die einzelnen Phasen im Lebenszyklus eines Registers gegliedert und weist in den folgenden Phasen Top-Level-Aufgaben und Anwendungsfälle Use Cases aus:

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

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

In aktueller Fassung werden über die verschiedenen Phasen 44 Top-Level-Aufgaben und 203 Use Cases abgebildet. Die Use Cases enthalten jeweils eine kurze Beschreibung der Aufgabe, benennen die notwendigen bzw. beteiligten Akteure sowie die Vor- und Nachbedingungen und führen einen Standardablauf auf. Anhand der Vorbedingungen werden die Abhängigkeiten zwischen den Anwendungsfällen ersichtlich, aus denen sich z. B. für die Planung eines Registers 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 zur Personal- und Kostenplanung für Kohorten und Register herangezogen und an das jeweilige Vorhaben angepasst 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, wie Rückmeldungen zu Präsentationen von Zwischenergebnisse zeigten. Abbildung 2 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.

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

– Abbildung 2 fortgeführt –

– Abbildung 2 fortgeführt –

Phase	Typ	Identifizier	Top-Level-Aufgabe/Use Case
Nutzung	S		
	TL	TL028	Studienunterstützung
	TL	TL029	Statistische Analyse
	TL	TL030	Berichterstattung
	TL	TL031	Organisation von Publikationen und Präsentationen
	TL	TL033	Datenintegration, Datenzusammenführung
Weiterentwicklung	S		
	TL	TL032	Weiterentwicklung Register/Kohorte
Abschluss	S		
	TL	TL034	Archivierung
	TL	TL035	Vernichtung der Daten, Anonymisierung
	TL	TL036	Close Out

Abbildung 2: KoRegIT – Liste der Top-Level-Aufgaben

### Nutzung des Anforderungskatalogs aus KoRegIT

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 Top-Level-Aufgaben, Use Cases und Phasen hinweg eine einheitliche Unterstützung am Rechnerarbeitsplatz für Mitarbeiter in ihren spezifischen Rollen angestrebt werden.

Die Abbildung von Kernprozessen durch IT-Komponenten zeigt beispielhaft eine Gegenüberstellung von Modulen der Toolbox aus dem Projekt Open Source Registry System for Rare Diseases in the EU (OSSE) [7] mit dem KoRegIT-Katalog.

1. OSSE Muster Dokumente
  - ▶ UC0019 Probandeninformation und Einwilligungserklärung erstellen
  - ▶ UC0015 Datenschutzkonzept erstellen
  - ▶ UC0041 Verfahrensanweisung erstellen
  - ▶ UC0050 Vertragsvorlagen erstellen
2. OSSE Formulareditor und Metadaten Repository
  - ▶ UC0034 Variablen, Ausprägungen und Integritätsbedingungen festlegen
  - ▶ UC0241 Datenbankschema erstellen
  - ▶ UC0203 Layout und Abfolge der Erfassungs- und Interviewformulare festlegen
  - ▶ UC0036 Erfassungsformulare erstellen
  - ▶ UC0037 Interviewformulare erstellen

3. OSSE Datenmanagement
  - ▶ UC0060 Berechtigungskonzept implementieren
  - ▶ UC0076 Nutzer, Rechte und Rollen verwalten
  - ▶ UC0077 Stammdaten verwalten
  - ▶ UC0088 Daten erheben
  - ▶ UC0092 Probandendatensatz bearbeiten
  - ▶ UC0093 Probandendatensatz deaktivieren/sperren
  - ▶ UC0094 Probandendatensatz endgültig löschen
  - ▶ UC0208 Probanden zu Kollektiv zuordnen
  - ▶ UC0242 Probandenberichte implementieren
  - ▶ UC0247 Datenbank versionieren
  - ▶ UC0162 Daten vernichten
  - ▶ UC0119 Daten exportieren und bereitstellen
4. OSSE ID Management über Mainzliste
  - ▶ UC0244 Zentrale Probandenliste führen
  - ▶ UC0090 Probanden registrieren
  - ▶ UC0163 Daten anonymisieren
5. OSSE Import Schnittstelle
  - ▶ UC0089 Daten importieren
  - ▶ UC0154 Datenintegration/Datenzusammenführung durchführen

Die OSSE-Toolbox bietet Funktionen beginnend bei einer metadaten-gestützten Daten- und Formulardefinition über ein webbasiertes Datenmanagement mit rollen-basiertem Zugriffskonzept bis hin zu Schnittstellen für die Vernetzung von OSSE-Registern. Sie unterstützt damit bereits 26 der 203 definierten Use Cases aus dem KoRegIT-Projekt, wobei fünf Tools bzw. Module identifiziert wurden. Geplant ist die Erweiterung des Funktionsumfangs und der Schnittstellen zur Unterstützung weiterer Use Cases (z. B. des Kontaktmanagements). Auf der Basis des KoRegIT-Katalogs kann die Realisierung eines Registers mit der OSSE-Toolbox systematisch geplant und umgesetzt werden.

## Aktueller Bestand

### Einleitung

Eine systematische Übersicht zu Registern, Kohorten und Data Repositories 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 kürzlich publiziert [9]. Hinweise auf große Kohorten lassen sich den Unterlagen der Nationalen Kohorte entnehmen [10].

Durch die somit fehlende Transparenz zu bestehenden Registern, Kohorten und Data Repositories 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 Initiative zur Implementierung eines Registerportals - eines Registers von Registern und Kohorten - gestartet [11]. Eine erste Datengrundlage hierfür sind die Ergebnisse aus Erhebungen zu Registern der Mitgliedsverbände der TMF, auf die im Weiteren ausführlicher eingegangen wird.

### **Erhebungen**

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 [12]. 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 [13]. 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 [2]. 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.

Von 25 Verbänden wurden 26 Register und Kohorten zurückgemeldet. Ein Register wurde von zwei Verbänden gemeldet. Die 26 Verbände des Nationalen Genomforschungsnetzes (NGFN Plus) waren einzeln um Rückmeldung gebeten worden. Bei keinem der 26 Verbände wurde jedoch ein Register betrieben, so dass das NGFN Plus in der Auswertung als ein Verbund geführt wird. Zusätzlich wurden fünf Register aus zwei früheren Mitgliedsverbänden gemeldet und in die Auswertung aufgenommen. Insgesamt liegen somit Angaben zu 31 Registern und Kohorten von 27 Verbänden vor (s. Tabelle 1). Damit betreiben etwa 30% der eingeschlossenen 88 Verbände ein Register oder eine Kohorte. Das erste Register wurde 1995 in Betrieb genommen, im Zeitraum zwischen 1995 und 2006 insgesamt 14 Register und zwischen 2008 und 2012 insgesamt 16 Register. Bei einem Register fehlt diese Information. Bei 26 Registern

werden die zentralen IT-Komponenten von Einrichtungen im universitären Umfeld betrieben. Bei fünf Registern finden sich andere Betreibermodelle.

Tabelle 1: Mitgliedsverbände der TMF mit Registern oder Kohorten

TMF-Mitglied	
▶ AID-Net	▶ Kompetenznetz Adipositas
▶ Centrum für Chronische Immundefizienz Freiburg	▶ Kompetenznetz Angeborene Herzfehler
▶ CURE-Net	▶ Kompetenznetz Asthma/COPD
▶ Deutsche Leberstiftung	▶ Kompetenznetz Degenerative Demenzen
▶ Deutsches Zentrum für Herz-Kreislauf-Forschung e. V.	▶ Kompetenznetz Multiple Sklerose
▶ Forschungsverbund Suszeptibilität: Hepatitis-C-Virus	▶ Kompetenznetz Parkinson
▶ MD-NET / Friedrich-Baur-Institut	▶ Kompetenznetz Vorhofflimmern
▶ GOLDnet	▶ Mukoviszidose Institut gGmbH
▶ IFB Adipositas Erkrankungen	▶ Netzwerk Ichthyosen
▶ IFB Zentrum für Schwindel	▶ PID-NET Konsortium
▶ INFOPAT	▶ SHIP-Studie
▶ IZKS Mainz	▶ TranSaRNet
▶ KKS Heidelberg	▶ ZKS Köln
	▶ ZKS Münster

*Aktuelle oder ehemalige Mitgliedsverbände der TMF, die in der Umfrage 2012 mindestens ein Register oder eine Kohorte gemeldet haben*

### Finanzierung und Ressourcen des Registers

Von 30 Registern liegen Angaben zu zentralen Ressourcen, von 12 Registern zu Ressourcen der Erhebungszentren vor. Zentral stehen zwischen 0,1 und 5,75 Vollkraftstellen für den Betrieb eines Registers zur Verfügung. Zwei Register werden nach Auslaufen der Förderung von Rekrutierung und Datenerhebung ohne zentrale Ressourcen gepflegt. Ein Register bedient sich eines externen Dienstleisters. Die

verbleibenden 27 Register verfügen im Median über eine Vollkraftstelle. Der Umfang des Studienpersonals wird von 10 Registern als nicht bekannt angegeben, acht Register verfügen über kein Studienpersonal in den Erhebungszentren, bei drei fehlen entsprechende Angaben. Damit verfügen 10 Register über Studienpersonal in den Erhebungszentren.

### Struktur des Registers

Alle Register halten ihre Daten zentral, bei einem Register ist dies mit einer dezentralen Datenhaltung kombiniert. Von 27 Registern liegen verwertbare Angaben zur Zahl der Erhebungszentren vor. Diese liegt zwischen 1 und 194; im Median beträgt die Anzahl der Erhebungszentren je Register 11. Folgende Einrichtungen beteiligen sich als Erhebungszentren: Universitätsklinik - 29 Register, nicht-universitäre Krankenhäuser - 20, niedergelassene Ärzte - 15, Probanden/Patienten - 7, Angehörige - 6, Sonstige - 6. Die geplante Laufzeit wird von 15 Registern mit „unbegrenzt“ angegeben, bei weiteren 11 Registern liegt diese bei mindestens 10 Jahren. Entsprechend ist bei den meisten Registern auch ein unbegrenzter Meldezeitraum vorgesehen.

### Probanden/Patienten im Register

Neun Register geben konkrete Zahlen zur Zielpopulation an, zwölf Register benennen die erwartete Fallzahl. Bei 14 weiteren Registern ist die Zahl der Teilnehmer unbegrenzt bzw. eine komplette Erhebung der Zielpopulation angestrebt. Von 27 Registern liegen Angaben zur Zahl aktuell registrierter Patienten vor. Im Median liegt diese Zahl bei 710 (Minimum 16, Maximum 143.000, Q1 287, Q3 8.728). Die Verteilung zeigt Abbildung 3.

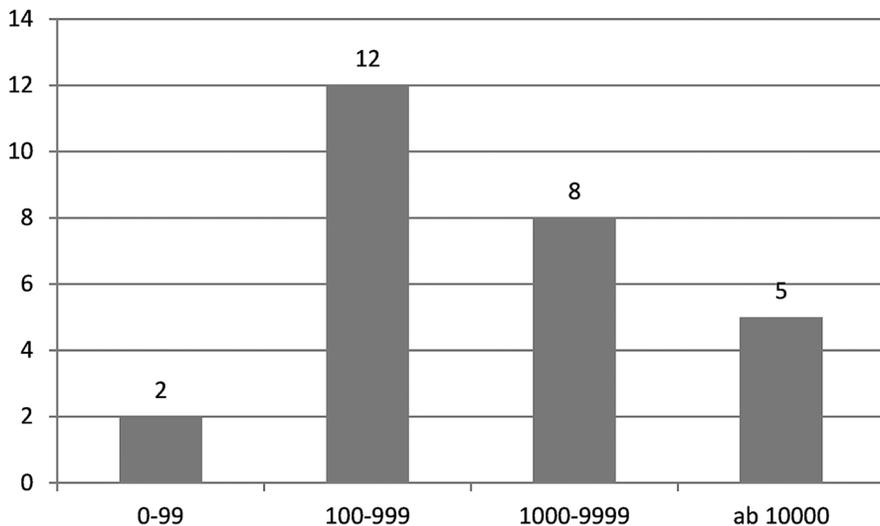


Abbildung 3: Zahl der Register nach Anzahl registrierter Patienten oder Probanden

## Merkmale im Register

Von 30 Registern liegen verwertbare Angaben zum minimalen und/oder maximalen Merkmalsumfang vor. Der minimale Merkmalsumfang liegt dabei zwischen 5 und 2700, der maximale zwischen 6 und 2700. Die Verteilung ist in Abbildung 4 dargestellt.

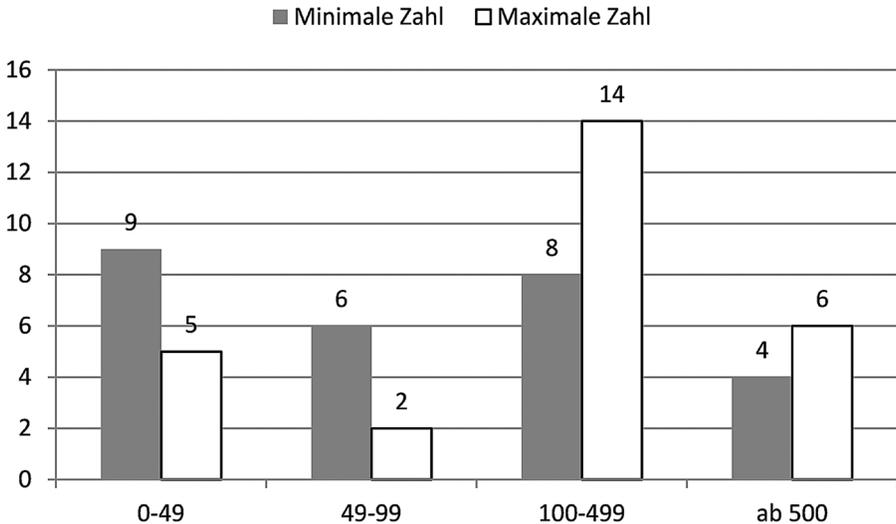


Abbildung 4: Anzahl der Register nach minimaler und maximaler Zahl der Merkmale

## Meldungen an das Register

Von allen Registern liegen Angaben zur Form der Meldungen vor. Bei 11 Registern erfolgen die Meldungen auf Papier, bei 13 Registern via Datensatz, bei 16 via Remote Data Entry. Bei 4 Registern werden Daten automatisch über Schnittstellen übernommen, hiervon ist einmal die Kopplung an Sekundärdaten, einmal die Übernahme von Stammdaten und einmal die Übernahme von Daten aus Studien angegeben. Ein Follow-up führen 28 Register durch; 3 Register verzichten auf ein Follow-up. Bei einem Register beträgt der Abstand zwischen den Meldungen 5 Jahre. Bei allen weiteren Registern liegt der Abstand bei maximal einem Jahr. Fünf Register geben ausdrücklich einen Abstand, zumindest zwischen der Erstmeldung und dem ersten Follow-up, von drei Monaten und kürzer an. Siebzehn Register führen ein Follow-up lebenslang durch. Bei ebenfalls 17 Registern erfolgt ein Record Linkage, 11 Register verneinen dies. Folgende Verfahren zum Record Linkage werden angegeben:

- ▶ Zusammenführung über eindeutiges Identifikationsmerkmal, Patientenliste o. ä. (11 Register)
- ▶ Neunstelliger alphanumerischer Code aufgedruckt auf einem „Patientenpass“ (1 Register)

- ▶ Abgleich mit Geburtsdatum und Initialen der Patienten o. ä. (2 Register)
- ▶ Deterministische und probabilistische Verfahren (1 Register)

### **Register-IT**

Dreizehn Register gaben eine Web- oder Internet-basierte Architektur an, 11 weitere eine Client-Server-Architektur. Vier Register machten ausschließlich Angaben zur Datenhaltung. Mit je einer Nennung waren es ASCII-Dateien, Microsoft Access und Oracle; ein weiteres Mal fehlte eine explizite Angabe. Vierundzwanzig Register benannten die Zahl der betriebenen Server. Neun Register betreiben einen Server, die höchste Zahl wird mit 10 angegeben. Der Median liegt bei zwei. Ab drei Servern lässt sich eine Verteilung von Applikation, Datenbank und Backup auf jeweils einen Server erkennen. Neun Register nutzen einen Backup-Server zur Ausfallsicherheit. Bei den Betriebskonzepten führt 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 ist ein kompletter Eigenbetrieb seltener.

Die zentralen Anwendungssysteme und -komponenten lassen sich bei 10 Registern dem Typ LAMP (Linux, Apache, MySQL, PHP) zuordnen, also der Nutzung von freier Software für Eigenentwicklungen. Acht Register setzen drei unterschiedliche kommerzielle Systeme zum Electronic Data Capture ein. Bei einem Register kommt ein klinisches Arbeitsplatzsystem aus dem Krankenhausbereich zum Einsatz. Von 13 Registern fehlen entsprechende Angaben oder die Angaben lassen sich nicht eindeutig zuordnen.

In Bezug auf die Unterstützung der Planung des Registers wurde viermal der Einsatz von Eigenentwicklungen zur Erstellung von Standard Operating Procedures angegeben. Vereinzelt wurden weiterhin Bürosoftware (Microsoft Excel, Visio) sowie Statistikpakete (SPSS, SAS) als Werkzeuge zur Planungsunterstützung genannt. Einundzwanzig Register nutzen Web-Browser als Clients für den Zugriff auf die zentralen IT-Komponenten. Bei einem Register handelt es sich um eine lokale Anwendung, einmal werden Oracle-Clients eingesetzt.

### **Dienstleistungen und Ergebnisse des Registers**

Bei 15 Registern erfolgt ein Abgleich mit weiteren Datenquellen. Dreiundzwanzig Register stellen den Mitgliedern des Verbunds, 14 der Öffentlichkeit regelmäßig Analysen zur Verfügung. Bei 15 Registern erhalten die Erhebungszentren die erfassten Daten zurück. Ebenfalls 15 Register geben die Daten an Dritte weiter. Eine Übersicht zur Erreichung der Ziele enthält Abbildung 5. Die folgenden sonstigen Ergebnisse wurden angegeben:

- ▶ Identifikation von Faktoren im Zusammenhang mit dem Erwerb/der Resistenz gegenüber einer akuten HCV-Infektion in einem Hochrisikokollektiv
- ▶ Untersuchung von Faktoren im Zusammenhang mit der Ausheilung der akuten Hepatitis-C-Infektion
- ▶ Meldung an das Landeskrebsregister
- ▶ Für klinische Routine nutzbare und praktikable systematische Datenhaltung
- ▶ Erfassung strahlentherapeutischer Spätfolgen
- ▶ Unterstützung externer Forscher
- ▶ Rekrutierungsübersichten
- ▶ Dosierungsempfehlungen für Psychopharmaka in der Kinder- und Jugendpsychiatrie
- ▶ Methodische Arbeiten

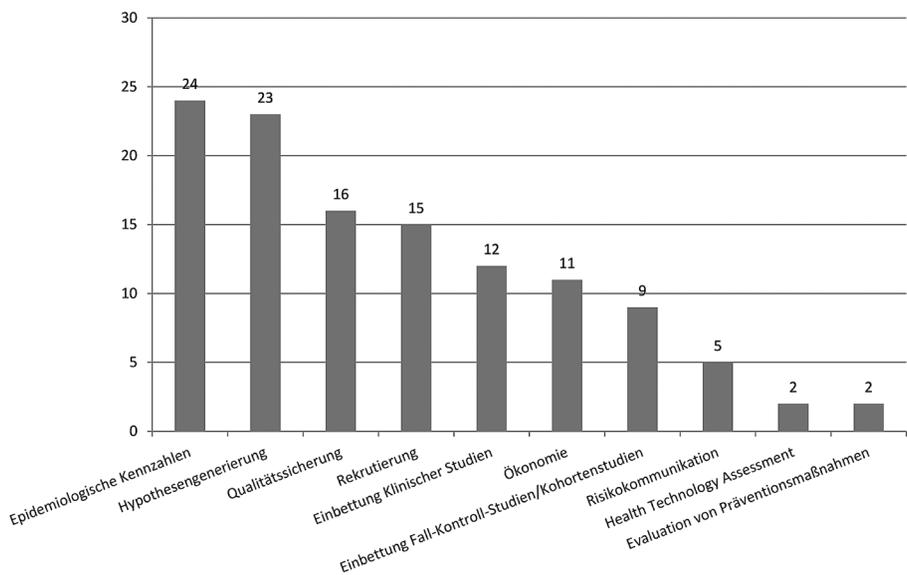


Abbildung 5: Von den Registern erreichte Ziele

### Verknüpfung mit Biomaterialbanken

Bei 9 Registern ist eine Biomaterialbank angeschlossen, bei 21 nicht. Einen Verweis auf Biomaterialien verwalten 17 Register, 13 nicht. Über weitergehende Informationen zu Biomaterialien verfügen 12 Register, 18 nicht.

### Bewertung und Handlungsbedarf

Die Erhebung unter den Mitgliedsverbänden der TMF von 2012 zeigte im Vergleich zu den vorangehenden Erhebungen in Bezug auf Register einen Mix aus Konstanz

und Wandel. Zu Kohorten und Data Repositories lassen sich bei deren geringer Zahl keine Aussagen treffen. Der Betrieb der IT-Komponenten liegt unverändert in der Hand universitärer oder Universitäts-naher Einrichtungen. Dennoch wird häufiger als bei den vorangehenden Erhebungen auf die Möglichkeit des Outsourcings zurückgegriffen. Bei rund der Hälfte der beschriebenen Betriebskonzepte wird zumindest teilweise diese Möglichkeit genutzt. Es dominiert dabei die Auslagerung von zentralen Hardwarekomponenten.

Die Erhebung bestätigt, dass kein Markt für kommerzielle Registersoftware existiert. Bei den kommerziellen Produkten dominieren EDC-Systeme aus dem Umfeld klinischer Studien, die in einem zum Teil langjährigen Prozess von den Anwendern an die Spezifika von Registern angepasst wurden. Führend sind jedoch Eigenentwicklungen unter Nutzung freier Software. Die Planung der Register findet weitgehend ohne spezifische IT-Unterstützung statt. Notwendig erscheint eine systematische Erhebung der von den bekannten kommerziellen und nicht kommerziellen Softwaresystemen unterstützten Funktionen für die Planung und den Betrieb von Registern und Kohorten. Mit dem dargestellten Anforderungskatalog zur IT-Unterstützung von Kohorten und Registern liegt ein hierfür passendes Raster vor.<sup>16</sup>

Zunehmende Aufmerksamkeit findet das Thema Datenqualität [14]. Für Register, Kohorten und Data Repositories liegt eine umfangreiche Empfehlung mit der Beschreibung von 51 Indikatoren zur Qualität von Daten vor [4]. 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 Tod. Die Zahl der Teilnehmer ist zumeist vierstellig. Der Wunsch nach möglichst vielen Informationen über die Teilnehmer zeigt sich in dreistelligen Zahlen von Merkmalen im Basisdatensatz. Gemessen an diesen Größenordnungen muss die Zahl von im Median nur einer Vollkraftstelle für den zentralen Betrieb des Registers als Unterfinanzierung bewertet werden. Register und Biomaterialbanken wachsen weiter zusammen. Achtzehn der 31 Register geben einen Bezug zu Biomaterialbanken an.

Bei den erreichten Ergebnissen deuten sich Verschiebungen an. Im Jahre 2009 dominierte noch die Unterstützung klinischer Studien durch eine Hypothesengenerierung, die Rekrutierungsunterstützung und die Kopplung des Registers mit klinischen Studien. Die Ermittlung epidemiologischer Kennzahlen und die Unterstützung der Qualitäts-

---

16 Die Notwendigkeit einer solchen Erhebung war auch das Fazit eines hierzu durchgeführten TMF-Workshops im Januar 2014 in Frankfurt (siehe [www.tmf-ev.de/news/1461](http://www.tmf-ev.de/news/1461)).

sicherung – also klassische eigenständige Ziele von Registern – wurden nun erstmals gleich häufig als Ergebnisse der befragten Register genannt.

Nach Kenntnis der Autoren liegt mit der für das IT-Review 2012/2013 der TMF durchgeführten Erhebung die umfangreichste Zusammenstellung über Register in Deutschland vor, von Krebsregistern abgesehen. Nach einem Aufbau vieler Register im Rahmen der Förderung von Kompetenznetzen in der Medizin hat 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.

Aus der Initiative von DNVF und TMF zum Aufbau eines Registerportals liegen erste Ergebnisse vor [15]. 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, Hinweise auf Dienstleistungen und entsprechende Regelungen sowie optional zum Download den Merkmalskatalog und das Datenschutzkonzept. Die entworfenen Verfahrensregeln gehen davon aus, dass die Nutzung des Registerportals für Interessenten wie für Vorhaben gebührenfrei ist. Die Verwaltung der Beschreibung von Registern und Kohorten würde weitestgehend an Vorhabensverantwortliche delegiert. Das Datenmanagement des Registerportals ist 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.

## Danksagung

Für die Unterstützung bei der Erhebung zu Registern danken wir Katharina Rimmele und Dr. Johannes Drepper von der TMF-Geschäftsstelle sowie Monika Darchinger vom IBE. Das Projekt „Von der Evaluierung zur Konsolidierung: Anforderungen an Kohortenstudien & Register IT (KoRegIT)“ wurde von der TMF unter der Nummer V078-01 gefördert. An KoRegIT sind das ZKS Köln, das IBE in München und das IMISE in Leipzig beteiligt. Besonderer Dank gilt den an der Erhebung teilnehmenden Mitgliedsverbänden der TMF.

## Literatur

1. Glicklich, R.E., Dreyer, N.A., Hrsg. *Registries for Evaluating Patients Outcomes: A User's guide*. 2 Aufl. 2010, Agency for Healthcare and Research Quality: Rockville, MD.
2. Müller, D., Augustin, M., Banik, N. et al., *Memorandum Register für die Versorgungsforschung*. Gesundheitswesen, 2010. **72**(11): S. 824–839.
3. DNEbM *Glossar zur Evidenzbasierten Medizin*. 2011. Deutsches Netzwerk Evidenzbasierte Medizin e.V., <http://www.ebm-netzwerk.de/was-ist-ebm/grundbegriffe/glossar> (Abruf: 2015-02-18).
4. Nonnemacher, M., Stausberg, J., Nasseh, D., *Datenqualität in der medizinischen Forschung. Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern*. 2 Aufl. 2014, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
5. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten – Generische Lösungen der TMF 2.0*. 2014, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
6. 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.
7. 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.
8. Michalik, C., Dress, J., Ngouongo, S.M. et al., *Requirements and tasks of cohorts and registers, the German KoRegIT project*. Stud Health Technol Inform, 2014. **205**: S. 1085–1089.
9. 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.
10. NatKo *The National Cohort. A prospective epidemiologic study resource for health and disease research in Germany*. 2011. Nationale Kohorte e.V., <http://www.nationale-kohorte.de/wissenschaftliches-konzept.html> (Abruf: 2015-02-18).
11. Stausberg, J., Semler, S.C., Neugebauer, E., *Ein Register von Registern und Kohorten: Das Registerportal von TMF und DNVF, 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. 484–486.

12. 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.
13. Stausberg, J., Altmann, A., Antony, G., Drepper, J., Sax, U., Schütt, A., *Register for networked medical research in Germany*. Applied Clinical Informatics, 2010. **1**(4): S. 408 - 418.
14. 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.
15. 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.



# Erhebung, Management und Verarbeitung digitaler Bilder

*Thomas Franke, Ulrich Sax*

## Anwendungsbereich

Die Bildgebung und zunehmend auch die 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 erfolgt der Einsatz computergestützter Verarbeitungsverfahren häufig hoch spezialisiert in eigenständigen Einrichtungen. In Ausnahmen wird die Bildverarbeitung in der klinischen Forschung parallel zur Routinebildgebung ohne zusätzliche Ressourcen durchgeführt. Im Folgenden werden die speziellen Anforderungen an die medizinische Bildverarbeitung strukturiert und im Überblick dargestellt. Die Strukturierung folgt dem vereinfachten, generischen Workflow in der forschungsbezogenen Bildverarbeitung (mit Studienbezug):

1. Formalisierung Studienprotokoll
  - 1.1 Demografische Daten
  - 1.2 Medizinische Daten
  - 1.3 Art der Bildgebung, betrachtete Region, Qualitätsparameter
  - 1.4 Festlegung Parameter Messung und Befundung (Konsensusverfahren)
2. Qualitätsmanagement
  - 2.1 Harmonisierung der Bildgebung: SOPs, Austausch von Phantomen
  - 2.2 Kontrolle der Bildgebung und der Einhaltung der SOPs
  - 2.3 Zertifizierung/Training der Reader (z. B. mittels ARCUS)
3. Rekrutierung der Probanden / Patienten
  - 3.1 Bevölkerungsbasiert
  - 3.2 Studienbasiert
4. Bildgebung
  - 4.1 DICOM<sup>17</sup>: CT, MRT, PET, Röntgen,
  - 4.2 Teilweise DICOM: Sonographie, Makroskopische Fotodaten (z.B. Dermatologie)
  - 4.3 Non-DICOM: Mikroskopische Daten (Digitale Pathologie)

---

17 Digital Imaging and Communications in Medicine: <http://medical.nema.org/standard.html>

5. Kommunikation der Bilddaten mit externen Einrichtungen
  - 5.1 Pseudonymisierung (DICOM Supplement 142)<sup>18</sup>
  - 5.2 CD/DVD-Versand
  - 5.3 Webupload per https oder sftp
  - 5.4 DICOM-Versand über spezielle Knoten
  - 5.5 Best practice: Guidelines for DICOM Media Exchange in Clinical Research<sup>19</sup>
6. Gegebenenfalls computergestützte Nachverarbeitung
  - 6.1 Registrierung, Segmentierung, Inhomogenitätskorrekturen, etc.
  - 6.2 Spezielle Software und Algorithmen
  - 6.3 Eigene Formate (z.B. NifTI, ANALYZE)
7. Messung
  - 7.1 Zertifizierte Ausstattung (Monitore, Umgebungsbedingungen)
  - 7.2 Spezielle Software
  - 7.3 Ausgebildete Fachkräfte
  - 7.4 Auftragsforschungseinrichtung
8. Auswertung der Messung
  - 8.1 Übergabe der Messwerte in unterschiedlichen Formaten
  - 8.2 Textuelle Befunde (Kategorisierung bzw. qualitative Auswertung)
  - 8.3 Statistische Auswertung von Messwerten z. B. in R, SAS, SPSS
9. Archivierung
  - 9.1 Beweissicherung
  - 9.2 Nachnutzung
10. Veröffentlichung
  - 10.1 Im Kontext einer Publikation
  - 10.2 Im Kontext eines Bild-Repositories

Durch Interviews mit mehreren Experten wurden drei generische Szenarien im Kontext der Bildverarbeitung in der klinischen Forschung identifiziert:

1. Erhebung der Bilddaten im Behandlungszusammenhang
2. Exklusive Gewinnung der Bilder im Forschungskontext
3. Bildgebung an Tieren und nicht-menschlichen Organismen

Bilddaten, die im Behandlungszusammenhang z. B. in klinischen Studien erhoben werden, werden im klinischen Picture Archiving and Communication System (PACS)

---

<sup>18</sup> [ftp://medical.nema.org/medical/dicom/final/sup142\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/sup142_ft.pdf)

<sup>19</sup> <http://swabik.offis.de/sites/swabik.offis.de/files/Guidelines.pdf>

abgelegt und verwaltet. Die entsprechenden Modalitäten (CT, MRT, PET, etc.) sind per DICOM direkt an das PACS angeschlossen. Die Befundung der Bilder findet an speziellen Befundarbeitsplätzen mit kalibrierten Monitoren und zertifizierter Software statt. Eventuell wird ein (pseudonymisierter) Export zur Weiterverarbeitung durchgeführt. Die exportierten Daten unterliegen dann nicht mehr der strukturierten Verwaltung durch das klinische PACS. Ihre Organisation und Archivierung muss gesondert vorgenommen werden.

Bilder, die ausschließlich für die Nutzung durch die Forschung gewonnen werden – oft auch an speziellen Forschungsmodalitäten – sind meist entkoppelt von der Verwaltung durch Abrechnungssysteme und Radiologie-Informationssysteme. Trotzdem werden sie häufig im klinischen PACS gespeichert, wenn kein spezielles Forschungs-PACS am Standort vorhanden ist. Die Daten werden mitunter direkt bei der Erhebung an der Modalität pseudonymisiert.

In der Pathologie sind bezogen auf die Menge insbesondere digitalisierte, histologische Schnitte und Sequenzierungsdaten hervorzuheben, bei welchen die Speicherung in proprietären Formaten und Systemen, bis hin zur Ablage im Filesystem, weit verbreitet ist.

Bei der Bildgebung an Tieren und nicht-menschlichen Organismen (z.B. Bakterien) ist die Verwaltung der Daten meist den Arbeitsgruppen überlassen. Die Bilder werden nicht in einem PACS, sondern häufig auf lokalen Workstations, File-Servern oder externen Medien gespeichert.

## **Betreiber, Anwender, Nutzer**

Üblicherweise wird am Standort eines klinischen Forschungsprojekts ein zentrales klinisches PACS für die Radiologie z. B. durch ein Rechenzentrum betrieben. An diesem PACS sollten alle bildgebenden Modalitäten und bildbeziehenden Abteilungen angeschlossen sein. Oftmals gibt es jedoch Ausnahmen bezüglich des Anschlusses einzelner Einrichtungen (z. B. bei der Anbindung der Pathologie, Zahnklinik, o.ä.).

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

Primäre Anwender eines Archivs für Forschungsbilddaten sind die klinischen Abteilungen, in denen Bilddaten zur Durchführung von klinischen Studien erhoben bzw. verarbeitet werden (Kardiologie, Radiologie, Nuklearmedizin, Pathologie), sowie Forschungsverbünde. Die Nutzer sind in erster Linie Ärzte/MTAs/Study Nurses, die die Daten erheben und in das System einstellen, sowie Forscher, die anhand der etablierten

Datenbasis eigene Auswertungen zur Beantwortung ihrer Forschungsfragestellungen 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 einigen Standorten gibt es separate Lehrbildsammlungen (z. B. für die Pathologie), für die eigenständige Infrastrukturen – abgekoppelt vom klinischen PACS bzw. den Forschungsarchiven – betrieben werden. Primäre Nutzer dieser Systeme sind Medizinstudenten und Dozenten.

## Rahmenbedingungen

Die technischen Rahmenbedingungen für die Erhebung, Kommunikation, Auswertung und Archivierung der Originalbilder werden durch den weitverbreiteten, konzeptbedingt jedoch sehr heterogenen DICOM-Standard geschaffen. Eine Ausnahme bildet noch die Digitale Pathologie: Diese wird zwar an einigen Standorten über Konvertierungs- bzw. Wrapping-Techniken DICOM-konform, an vielen anderen jedoch proprietär gelöst. Das für diesen Zweck entwickelte Supplement 145<sup>20</sup> des DICOM-Standards findet in der Praxis noch so gut wie keine Anwendung, wobei die Digitalisierung selbst hier noch immer das größte Hemmnis für den Einsatz darstellt (vergl. [1]). Bei der Nachverarbeitung und der – nicht rein visuellen – Auswertung der Bilddaten für die Forschung sind die Rahmenbedingungen nicht so klar vorgegeben: Es gibt diverse Softwaretools zur Verarbeitung, spezielle Datenformate und keine einheitlichen Kommunikationsprotokolle bzw. Schnittstellen zu den Archivierungssystemen. Im Rahmen der Aktivitäten der Standardisierungsorganisation IHE<sup>21</sup>, sowie auch aus den Fachdisziplinen selbst (Bsp. Pathologie, vergl. [1]) gibt es jedoch Bestrebungen, auch diese Berichte weiter zu standardisieren.

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 (z. B. Zertifizierung von Software) und die Regeln zur guten wissenschaftlichen Praxis (Aufbewahrungsfristen für Daten, die Grundlage wissenschaftlicher Veröffentlichungen sind) vorgegeben. Im Kontext klinischer Studien müssen zusätzlich

---

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

21 <http://ihe.net>

internationale Richtlinien, wie z. B. CFR 21 Part 11 der US Food and Drug Admision (FDA)<sup>22</sup>, sowie die Good Clinical Practice (GCP) der ICH<sup>23</sup> berücksichtigt werden.

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 seitens des Förderers möglich). An dieser Stelle sei jedoch darauf hingewiesen, dass die notwendige Betriebsdauer der Infrastruktur, vor allem mit Blick auf gesetzliche Aufbewahrungsfristen und einer für die wissenschaftliche Nachnutzbarkeit gewünschten Langzeitarchivierung, die eigentliche Projektlaufzeit übersteigen kann.

## Kategorisierung der IT-Landschaft

Die betrachteten IT-Systeme zur Bildverarbeitung können grob in zwei Kategorien unterschieden werden:

1. Applikationen aus der Krankenversorgung, die Daten für den Forschungseinsatz exportieren.
2. Systeme, die speziell für den Einsatz in der Forschung beschafft und betrieben werden.

In die erste Kategorie fallen RIS und PACS, die mittlerweile zum Standardrepertoire jedes Universitätsklinikums gehören, sowie diverse bildgebende Modalitäten (CT, MRT, PET, Sonographie, etc.). Alle Bildgebungs- und Bildbetrachtungsmodalitäten, die im Rahmen der Routineversorgung von Patienten genutzt werden, sind im Regelfall an ein PACS angeschlossen.

Die Kommunikation nach „Extern“ aus den Versorgungssystemen heraus findet in der Regel über Exporte aus dem PACS statt. Diese Exporte lassen sich auf verschiedene Arten umsetzen: über (1) CD/DVDs, über (2) Webuploads – nur bei kleineren bis mittleren Datenaufkommen geeignet – oder (3) über spezielle DICOM Knoten/Gateways. In vielen Fällen befindet sich das PACS jedoch in einem Netzsegment, aus dem

---

22 <http://www.fda.gov>

23 <http://www.ich.org>

ein direkter Versand in die „Außenwelt“ aus Sicherheitsgründen nicht möglich ist, sodass zunächst eine Lücke überwunden werden muss, bevor ein Webupload durchgeführt werden kann. Die pragmatische, wenngleich auch etwas umständliche Lösung ist oftmals die Daten zunächst aus dem PACS auf CD zu brennen und sie anschließend an einem PC mit Internetzugang einzulesen und hochzuladen.

In die zweite Kategorie fallen Modalitäten, die für den reinen Forschungseinsatz beschafft worden sind (z. B. Präparate-Scanner oder ein 7 Tesla Forschungs-MRT), sowie Forschungsbilddatenarchive und Kollaborationssysteme. Des Weiteren gehören verschiedene Verarbeitungs- und Auswertungs-Tools zu dieser Produktkategorie. Forschungskomponenten, die nur an einem Standort genutzt werden, sind häufig enger mit den vorhandenen Versorgungssystemen verzahnt, als Lösungen für Verbundforschungsprojekte. Für letztere werden oftmals eigene Systeme – losgelöst von der Standortinfrastruktur – betrieben.

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 etc. über gesonderte Verträge sicherzustellen. Die Rentabilität einer solchen Auslagerung hängt stark von den Gegebenheiten des Einzelfalls ab, wobei im Falle der Pathologie mit beachtlichen Datenmengen zu rechnen ist.

## Aktueller Bestand und Erhebungsergebnisse

Die aktuelle Erhebung versucht einen Überblick über die verfügbaren Software-Produkte zu geben. Aufgrund einer Vielzahl von Open-Source-Softwarestacks und einer sehr starken Spezialisierung auf einzelne Fachdisziplinen bei den Tools zur Bildverarbeitung haben die Ergebnisse keinen Anspruch auf Vollständigkeit. Aus denselben Gründen wurde auch von einer umfangreichen Analyse der einzelnen Produkte abgesehen. Die folgenden Software-Produkte wurden identifiziert:

1. Systeme zur Verwaltung, Archivierung und zum Austausch von Bildern
  - ▶ XNAT<sup>24</sup> (Washington University of St.Louis, MO, USA) ist eine 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 Analyze-Format und mit allgemeinen Daten im XML-Format umgehen.

---

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

- ▶ Chili PACS<sup>25</sup> und die webbasierte Telemedizinakte der Chili GmbH Heidelberg/Dossenheim zum institutionsübergreifenden Austausch von Bilddaten z.B. in verschiedenen Kompetenznetzen in der Medizin, aber auch für den Bildaustausch in verschiedenen Berufsgesellschaften (s. DGU)
- ▶ MDPE<sup>26</sup> – 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.
- ▶ The Common Toolkit (CTK<sup>27</sup>). CTK ist eine Open-Source DICOM Plattform, mit der es möglich ist, DICOM Bilder aller Art zu speichern, anzuzeigen, zu senden und zu empfangen.
- ▶ Suite des Fraunhofer-Instituts für Computerunterstützung in der bildbasierten Medizin MEVIS<sup>28</sup> in Bremen
- ▶ TRIAL-Connect<sup>29</sup> der Telekom Healthcare Solutions. Eine webbasierte Forschungsplattform zur zentralen Verwaltung und Archivierung von Bilddaten in klinischen Studien.
- ▶ OMERO<sup>30</sup> des OME (Open Microscopy Environment) Consortium. 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).

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

- ▶ Freie und Open-Source-Software: MITK<sup>31</sup>, FSL<sup>32</sup>, SPM<sup>33</sup>, FreeSurfer<sup>34</sup>, 3DSlicer<sup>35</sup>, Fiji<sup>36</sup>, Bio-Formats<sup>37</sup>, OMERO<sup>32</sup>

---

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

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

27 <http://www.commontk.org>

28 <http://www.mevis.fraunhofer.de>

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

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

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

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

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

34 <http://freesurfer.net/>

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

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

37 <http://loci.wisc.edu/software/bio-formats>

- ▶ Kommerzielle Produkte: BrainVoyager<sup>38</sup>, CMR42<sup>39</sup>, Definiens<sup>40</sup>
3. Produkte, die in der Lehre eingesetzt werden
- ▶ Surgery-Planner von MEVIS/Gruppe Prof. Preim
  - ▶ Digital SlideBox von Leica
  - ▶ Digitale Pathologielösungen (z.B. an den Universitätskliniken Magdeburg und Heidelberg)
  - ▶ MedinfoGrid-Projekt: MIRC<sup>41</sup> (RSNA)

## 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. Stattdessen ist es vorzuziehen, auf ein bestehendes System zu setzen und dieses an die eigenen Bedürfnisse anzupassen. 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 Workflows und des Qualitätsmanagements abhängig gemacht werden.

Perspektivisch sollte die technologische Weiterentwicklung der Bildgebung in der Krankenversorgung und der Forschung besser als bisher verzahnt werden, auch wenn unterschiedliche Innovationszyklen immer für ein divergentes Moment sorgen werden. Allerdings sehen die Autoren eine engere Abstimmung der verschiedenen Akteure (z. B. Fachgesellschaften, Standardisierungsorganisationen) als notwendig an, um Synergieeffekte besser nutzen zu können. Insbesondere die einrichtungübergreifende Kommunikation von Bilddaten, wie sie in Forschungsverbänden üblich ist, wird oft noch proprietär gelöst, obwohl IHE mit XDS.b und dessen Erweiterung XDS-I entsprechende Profile anbietet. Deshalb sollte geprüft werden, ob diese Profile und Konzepte auch im Rahmen der Forschung angewendet werden können. Dabei ist auch die Notwendigkeit einer standardisierten Anbindung von ID-Managementfunktionen, z. B. zur Pseudonymisierung von Bilddaten, zu berücksichtigen (s. auch das Kapitel zum Identitätsmanagement im vorliegenden Bericht).

---

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

39 <http://www.circleevi.com/>

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

41 <http://mirc.rsna.org>

Eine weitere Standardisierung über das bisherige Anwendungsspektrum von DICOM hinaus, welches vorwiegend Kommunikationsaspekte abbildet, erscheint dringend notwendig. Insbesondere für die Nachverarbeitung von Bilddaten sowie die Dokumentation von Verarbeitungsschritten und Ergebnisdaten gibt es bisher keine Standards, was auch daran liegen könnte, dass die Bildnachverarbeitung bisher kaum in den klinischen Workflow integriert ist. Automatische Segmentierungen werden beispielsweise bisher nur in wenigen Bereichen angewendet. Dies könnte u. a. auch an dem Fehlen einer anerkannten Plattform zur Verbreitung neuer, standardisierter und qualitätsgesicherter Bildverarbeitungsalgorithmen und Tools liegen.

Die Digitale Pathologie wird hohe Anforderungen an die IT-Infrastruktur und insbesondere an das PACS stellen. Zum einen sind die Bilddaten mit Dimensionen jenseits der 100.000 Zeilen oder Spalten wesentlich größer, als es der DICOM-Standard bisher vorsah, zum anderen wird eine verlustbehaftete Kompression für diese Bilder zwingend erforderlich sein. Dies bedeutet, dass jede unnütze Rekompensation vermieden werden muss, da diese einen Informationsverlust zur Folge haben kann. Außerdem können Bilddaten nur bedingt, wie bisher üblich, auf Verdacht zu verschiedenen Befundarbeitsplätzen übertragen werden. Vielmehr ist die Bildverteilung aus dem Archiv mittels eines Streamingverfahrens sinnvoll. Grundsätzlich ist hierbei jedoch festzuhalten, dass bisher keiner der großen PACS-Hersteller eine hinreichende Integration der Digitalen Pathologie vorweisen kann. JPEG2000, als offenes Format, bietet hier aktuell einige Vorteile im Hinblick auf die Zukunftssicherheit, da es alle DICOM-Bildarten effizient verarbeiten kann, schon als zusätzlich Transfersyntax in den DICOM-Standard integriert wurde, die progressive Speicherung und Wiedergabe von Auflösungspyramiden unterstützt und zudem ähnliche Bilder einer Serie zu Stapeldateien zusammen fassen kann. Weiterhin ist das speziell für JPEG2000 angepasste Streamingverfahren JPIP offizieller Bestandteil der Definition und ebenfalls in den DICOM-Standard aufgenommen. Allerdings ist JPEG2000 hinsichtlich der Zugriffsgeschwindigkeit den proprietären Formaten häufig unterlegen.

Der oft hohen Interobserver-Variabilität bei visuellen Auswertungen sollte durch entsprechende Reader-Schulungen, Zertifizierungen und ein adäquates Qualitätsmanagement begegnet werden. Auch hierfür wird eine passende IT-Unterstützung benötigt, z. B. in Form webbasierter Plattformen mit einer geeigneten Referenzdatenbank. Ein vielversprechender Ansatz ist die im Rahmen der SHIP-Kohorte (vergl. [2]) entwickelte ARCUS-Plattform<sup>42</sup>.

---

42 Die ARCUS-Plattform wurde in der Sitzung der AG IT-Infrastruktur und Qualitätsmanagement der TMF am 20.11.2012 vorgestellt. Eine Veröffentlichung soll folgen.

## Danksagung

Wir danken

- ▶ Prof. Dr. Dr. Johannes Bernarding, Magdeburg
- ▶ Prof. Dr. Hans-Peter Meinzer, Heidelberg
- ▶ Prof. Dr. Peter Mildenerger, Mainz
- ▶ Prof. Dr. Peter Hufnagl, Berlin
- ▶ Dr. Ralf Zwönitzer, imassense Deutschland GmbH, Berlin
- ▶ Dipl.-Inf. Norman Zerbe, Berlin

für ihren fachlichen Rat und ihre Mithilfe bei der Erhebung der gegenwärtigen IT-Landschaft.

## Weiterführende Literatur

1. Hufnagl, P., Zerbe, N., Schlüns, K., *Virtuelle Mikroskopie in der onkologischen Diagnostik*. *Onkologie*, 2012. **18**: S. 409–418.
2. Völzke, H., Schmidt, C.O., Hegenscheid, K. et al., *Population Imaging as Valuable Tool for Personalized Medicine*. *Clinical Pharmacology and Therapeutics*, 2012. **92**(4): S. 422–424.
3. Rojo, M.G., Daniel, C., Schrader, T., *Standardization efforts of digital pathology in Europe*. *Anal Cell Pathol (Amst)*, 2012. **35**(1): S. 19–23.
4. Channin, D.S., Mongkolwat, P., Kleper, V., Sepukar, K., Rubin, D.L., *The caBIG annotation and image Markup project*. *Journal of Digital Imaging*, 2010. **23**(2): S. 217–225.
5. Engelmann, U., *Teleradiologie: Der Weg von der Forschung in die Regelversorgung*. *mdi-Forum der Medizin\_Dokumentation und Medizin\_Informatik*, 2011. 2: S. 72–79.
6. IHE Cookbook: Einrichtungsübergreifende elektronische Bild- und Befundkommunikation. 2012. IHE Deutschland e.V., <http://www.ihe-d.de/index.php/downloads/category/3-cookbook> (Abruf: 2013-04-18).
7. Marcus, D.S., Erickson, B.J., Pan, T., *Imaging Infrastructure for Research. Part 2. Data Management Practices*. *Journal of Digital Imaging*, 2012. **25**(5): S. 566–569.
8. 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.
9. Ruchholtz, S., Mand, C., Lewan, F., Dankowski, C., AKUT-Steering-Committee, Kühne, C., Siebert, H., *Regionalisation of trauma care in Germany: the "TraumaNetwork DGU-Project"*. *Eur J Trauma Emerg Surg*, 2011. **38**: S. 11–17.

# Biobanken

*Sara Y. Nußbeck, Andreas Beck, Daniela Skrowny, 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 investiert das Bundesministerium für Bildung und Forschung (BMBF) von 2011 bis 2016 über 20 Millionen Euro.<sup>43</sup>

In diesem Kapitel über IT-Systeme zur Unterstützung des Biobanking (Biobankmanagementsysteme) sind nur 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 die Empfehlungen der TMF-Datenschutzkonzepte [4] zur informationellen Gewaltenteilung zurückzuführen, die allerdings nicht zwangsweise die Nutzung unterschiedlicher Softwaresysteme vorschreiben. 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. 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 prä-analytischer 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.

---

43 siehe <http://www.gesundheitsforschung-bmbf.de/de/2638.php>

Obwohl im Bereich der Biobanken meist zwischen „liquid biobanking“ und „solid biobanking“ unterschieden wird, ist diese Unterscheidung im Bereich der Softwareprodukte eher irrelevant, da sich lediglich der Workflow und die 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. Bereits 2009 wurde im Rahmen des IT-Strategie-Projektes der TMF ein Teilprojekt zum Thema „IT-Werkzeuge für Biomaterialbanken“ durchgeführt. Damals wurde ein Anforderungskatalog erstellt [3], mit dessen Hilfe eine Marktanalyse von 21 auf dem Markt erhältlichen bzw. in größeren Biobanken eingesetzten Softwarelösungen für das Management von Biobanken durchgeführt wurde. In Bezug auf die kommerziell angebotenen Produkte wurden deren Verfügbarkeit sowie der technische Produktsupport auf dem deutschen Markt als wesentliches Kriterium definiert. Zum damaligen Zeitpunkt waren auf dem deutschen Markt Vertriebspartner bzw. Vertriebsniederlassungen lediglich für die Produkte Cresalys (Alphamatrix), Starlims (Axel Semrau), Nautilus (ThermoFisher Scientific) und auch Sapphire (Labvantage) gegeben. Unmittelbaren Systemsupport gab es damals in Deutschland lediglich für Starlims und Nautilus. Des Weiteren wurde der Einsatz entsprechender IT-Lösungen in den großen deutschen Biobanken untersucht. Dabei zeigte sich, dass bis 2009 fast ausschließlich projektspezifische Eigenentwicklungen genutzt wurden.

Fazit: Die Dokumentation und Verwaltung von Biomaterialien erfordert heute eine professionelle IT-Infrastruktur. Sowohl die Vielfalt der Softwarelösungen als auch die Komplexität des Aufbaus, Betriebs und der Vernetzung von Biobanken stellen viele Forschungsverbünde vor große Herausforderungen. 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 [8-10] dazu beigetragen, dass Forschungsverbünde auf diesem Wissen aufbauen können.

## Betreiber, Anwender, Nutzer

Was ist die Definition von „Biobank“? Eine Online-Umfrage unter Vertretern der internationalen Biobankengesellschaft ISBER führte zu dem Konsens, dass es sich bei Biobanken um biologische Sammlungen von Menschen, Tieren, Pflanzen oder Mikroben handeln kann, dass assoziierte Daten dazu gehören und dass diese Sammlungen nach professionellen Standards verwaltet werden müssen [11]. Bezüglich der Größe, des Zwecks oder der Zugangsregelung gab es keine Einigkeit. Es ist daher nicht verwunderlich, dass, je nach organisatorischem Aufbau und Größe der Biobank, die Betreiber, Anwender und Nutzer eines Biobankmanagementsystems stark variieren können.

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.

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, 12]. Insgesamt stellt sich in einem Verbund immer die Frage einer alle „Teilbiobanken“ übergreifenden Abfrage/Retrievalmöglichkeit von Proben. Für diesen Zweck 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 können unabhängig davon zum jeweiligen Proben-/Lagerortmanagement weiter verwendet werden.

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. [10] verwiesen. Grundsätze für eine datenschutzgerechte IT-Unterstützung im Biobanking sind in der Regel die Pseudonymisierung personenbezogener Daten, eine Trennung der Datenhaltung für Probandendaten und Probandendaten sowie eine Trennung der Datenhoheit und Verantwortlichkeiten. Da Biomaterial meistens genetische Informationen enthält, die unter Umständen einen Rückschluss auf die Identität und 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 [13] und im neuen 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.

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 [14]
- ▶ 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) [15]
- ▶ 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 [16]
- ▶ OECD Best Practice Guidelines for Biological Resource Centres, veröffentlicht von der Organisation for Economic Co-operation and Development (OECD) [17]
- ▶ OECD Guidelines on Human Biobanks and Genetic Research Databases, ebenfalls veröffentlicht von der OECD [18]
- ▶ Die Empfehlung Rec (2006) 4 on research on biological materials of human origin des Ministerkomitees des Europarates [19]
- ▶ Die Deklaration des Weltärztebundes zu Ethischen Grundsätzen für die medizinische Forschung am Menschen, „Declaration of Helsinki“ [20]

Im Zuge der wachsenden Bedeutung von Biobanken für die Forschung wird seit einigen Jahren auch intensiv über die Notwendigkeit eines Biobankengesetzes oder eines gesetzlich verankerten Biobankgeheimnisses, wie vom Deutschen Ethikrat vorgeschlagen [14], diskutiert<sup>44</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. 21].

**Organisatorisch:** Organisatorisch können Biobanken gemäß einer der drei von Pommerening [9] 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) 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 Biobankmanagementsystem zu etablieren, in dem sowohl eine zentrale Biobankstruktur als auch dezentral verteilte Biobanken abgebildet werden können. Da entsprechende Systeme an vielen deutschen Universitätsklinikum 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 des 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 bei Metadaten zur Charakterisierung von Biobanken und auch den Phänotypdaten auf entsprechende Standards zu setzen. Beispielhaft seien hier SPREC [5, 6], BRISQ [22], MIABIS [23] sowie der ADT-Datensatz in den Comprehensive Cancer Center genannt. Aktuell widmet sich die Arbeitsgruppe

---

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

ESBBperanto der Europäischen Biobankengesellschaft (ESBB) generisch dem Thema des Datenaustauschs zwischen Biobanken.

## 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.

Interessant ist jedoch zu sehen, dass die meisten Produkte zum Management von Biobanken aus dem Bereich der Labor-Informationen-Management-Systeme (LIMS) oder aus dem Anwendungsfeld der Pathologie stammen. Darüber hinaus gab es in der Anfangszeit Speziallösungen, wie z. B. die Biobank der Medizinischen Universität Graz, die auf einer Software für Museen basierte.

Der wichtigste Aspekt einer Biobank ist die Verknüpfung der Biomaterialproben mit ihren Annotationsdaten. Zu den Annotationsdaten gehören vorrangig die phänotypischen und klinischen Verlaufsdaten (inkl. Bilddaten) der Probenspender. Es können aber auch Ergebnisse aus molekularen oder genetischen Analysen sein.

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 gesamte IT-Infrastruktur einer Biobank umfasst zur Verknüpfung der Proben mit strukturierten Phänotypdaten in der Regel auch das Clinical Data Warehouse (CDW), das Röntgeninformationssystem (RIS) und das Picture Archiving and Communication System (PACS). 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 lediglich 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. [24] beschrieben.

## Aktueller Stand

Die im Weiteren aufgeführten Angaben zum aktuellen Stand des Einsatzes von Biomaterialverwaltungssoftware beruhen auf drei unterschiedlichen Informationsrecherchen. Für eine Aussage zur Verbreitung von Biomaterialverwaltungssoftware in Deutschland wurden die vom BMBF geförderten zentralisierten Biobanken (cBMBs), die von der Deutschen Krebshilfe geförderten Comprehensive Cancer Center Biobanken (CCCs) und die vom BMBF geförderten deutschen Zentren für Gesundheitsforschung befragt (wobei berücksichtigt werden muss, dass sich dadurch Befragungen teilweise überlappen, da die Biobanken universitärer Standorte teilweise allen drei Kategorien zugeordnet werden können). Darüber hinaus wurde im Rahmen der Sitzung der AG Biobanken der TMF im September 2014 in Mannheim und im Anschluss über den E-Mail-Verteiler der AG Biobanken und des Deutschen Biobankenregisters ein kurzer Fragebogen zur Verwendung von Biomaterialverwaltungssoftware ausgeteilt bzw. versandt. Die Angaben zum internationalen Stand beruhen auf einer Online-Umfrage der ISBER Working Group (WG) Informatics von 2012 [25], an der sich 71 Personen (alle ISBER Mitglieder) beteiligten. Ausschnitte dieser Umfrage werden nachfolgend erläutert.

Nicht eingegangen wird im Folgenden auf Open-Source-Lösungen wie z. B. OBiBa oder Bika LIMS, die in den Umfragen nicht auftauchen bzw. in der deutschen Community derzeit offensichtlich auch nicht oder kaum eingesetzt werden. Aufgrund fehlenden Feedbacks zu diesen Softwaresystemen ist auch jenseits möglicher allgemeiner Vorbehalte gegenüber Open-Source-Software keine Einschätzung möglich, warum diese in diesem Anwendungsfeld so wenig Verbreitung finden.

## Deutschland

Während 2009 in den deutschen Biobanken kaum kommerzielle Biobankmanagementsysteme eingesetzt wurden, hat sich die Situation bis Ende 2014 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 Instituten oder projektbasierten Biobanken. Insgesamt gaben knapp 60% der befragten Biobanken an, für die Verwaltung der Biomaterialien auf ein kommerzielles, professionelles Biobankmanagementsystem zu setzen (Abbildung 6). Von den 36%, die bisher nicht auf ein solches kommerzielles Biobankmanagementsystem, sondern auf Eigenlösungen oder adaptierte Softwarelösungen setzen, ist nur bei einem Drittel ein Wechsel zu einem kommerziellen Produkt geplant.

### Benutzt die Biobank eine Software zur Verwaltung des Biomaterials (n=30)?

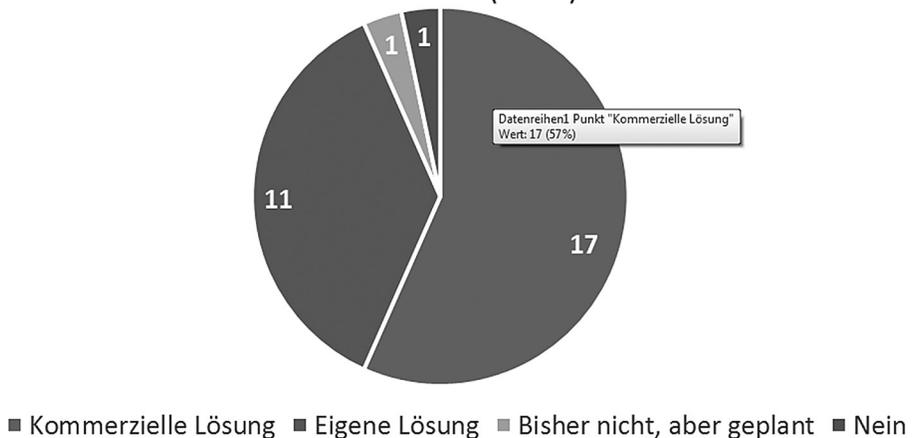


Abbildung 6: Ergebnis der Umfrage unter deutschen Biobanken im Oktober 2014

Auch wenn einzelne Standorte mit eigenentwickelten Lösungen arbeiten (z. B. die Biomaterialbank Nord am Forschungszentrum Borstel), so haben sich doch die meisten universitären Standorte mittlerweile für die Einführung eines kommerziellen Biobankmanagementsystems entschieden bzw. nutzen ein solches schon (50% bereits in der Produktivphase). Zwei Produkte scheinen sich dabei in Deutschland als Marktführer zu etablieren. Es sind dies die Systeme Starlims (Abbott Informatics Germany GmbH) und CentraXX Biobank (KAIROS GmbH). Ein wichtiger Anstoß zu dieser Entwicklung war die Präsentation, Verbreitung und Diskussion des Abschlussberichts des IT-Strategie-Projekts (2009) der TMF zur IT-Unterstützung von Biobanken in der AG

Biomaterialbanken. Von den durch das BMBF geförderten fünf cBMBs haben sich die Standorte Berlin, Kiel und Würzburg für das System CentraXX Biobank der Kairos GmbH entschieden, während die Biobanken in Heidelberg und Aachen auf Starlims der Firma Abbott setzen. CentraXX Biobank befindet sich derzeit an den genannten Standorten in der Anpassungsphase bzw. teilweise schon im Produktivbetrieb. Zu den weiteren relativ neuen Kunden von CentraXX gehören Frankfurt (UCT), das DKTK (wobei hier in einzelnen Zentren derzeit vor allem die klinische Annotation über einen sogenannten DKTK-Brückenkopf vermittelt wird, nicht aber die eigentliche Bioprobenverwaltung mit CentraXX erfolgt), die Nationale Kohorte, zentrale Biobank und CCC in Lübeck und die Universitäten Tübingen und Freiburg. Die beiden Standorte mit Starlims haben sich bereits zu einem früheren Zeitpunkt entschieden, so dass die Software dort bereits seit längerem produktiv eingesetzt wird. Weitere Universitätsklinika (Bochum, Göttingen, Jena, Erlangen, Münster, Deutsches Herzzentrum Berlin, Gießen & Marburg) sowie Roche Diagnostic in Penzberg haben sich ebenfalls für Starlims entschieden bzw. nutzen dieses bereits produktiv.

Bei den durch die Deutsche Krebshilfe geförderten Biobanken der Comprehensive Cancer Center stellt sich die Lage noch recht gemischt dar. Einige der Standorte nutzen weiterhin lokale Eigenentwicklungen (Dresden, Hamburg, Köln/Bonn), andere hingegen haben sich mittlerweile für kommerzielle Lösungen entschieden, die teilweise schon länger im Einsatz sind (Starlims in Erlangen und Heidelberg, Nautilus in Ulm) bzw. bei denen in den letzten zwei Jahren mit der Einführung begonnen wurde (CentraXX in Berlin, Frankfurt und Tübingen, eBioControl der Fa. GEFAT-IT in Essen).

Bei den Deutschen Zentren der Gesundheitsforschung ist die Analyse aufgrund der heterogenen Zusammensetzung und den vielen beteiligten Standorten etwas schwieriger. Alleine das Deutsche Zentrum für Herz-Kreislauf-Forschung (DZHK) besteht aus 26 Einrichtungen in sieben Zentren. Generell lässt sich die Infrastruktur der DZGs einer der folgenden drei Kategorien zuordnen: A) Nutzung des jeweils lokalen Biobankmanagementsystems der einzelnen Standorte mit lokalem Verbleib der Proben und übergreifender Zusammenführung von Proben- und Phänotypdaten, B) zentrale Software zur Verwaltung der Materialien an den einzelnen Standorten und C) zentrale Biobank mit zentralem Biobankmanagementsystem. Teilweise kann man die DZGs auch mehreren Kategorien zuordnen.

Ein Beispiel liefert das DZHK: Hier ist geplant, dass Standorte mit vorhandener professioneller Software diese weiterverwenden können und Schnittstellen geschaffen werden. Sollte ein Standort bisher keine eigene Lösung haben, kann die übergreifende Softwarelösung zur Verwaltung des Biomaterials genutzt werden. Das Deutsche Konsortium für Translationale Krebsforschung (DKTK) setzt derzeit für alle bisher schon erhobenen Daten der einzelnen Zentren auf ein föderiertes „Brückenkopf-Konzept“ mit

dezentralen Suchen und Freigabemechanismen. Für die zukünftig mit entsprechender Patienteneinwilligung erhobenen Daten ist – zumindest für einen gemeinsam definierten Kerndatensatz – eine übergreifende Zusammenführung der Datenbestände gemäß Kategorie A geplant. Des Weiteren zählt das Deutsche Zentrum für Infektionskrankheiten (DZIF) mit drei dezentralen Biobanken mit unterschiedlichen Schwerpunkten und einer übergreifenden Webseite zur Probenverfügbarkeit dazu, wie auch das Deutsche Zentrum für Lungenkrankheiten (DZL) mit einer Datenzusammenführung in Kooperation mit dem Deutschen Biobanken-Register. Bisher hat sich noch keines der Zentren für einen reinen Ansatz nach Kategorie B) entschieden. Zur Kategorie C) gehören das Deutsche Zentrum für Diabetesforschung (DZD) und das Deutsche Zentrum für Neurodegenerative Erkrankungen (DZNE). Bei beiden soll eine zentrale Biobank mit einem zentralen Proben- und Datenmanagement etabliert werden. Die Frage nach den passenden Softwarelösungen für die unterschiedlichen Ansätze ist bei den DZGs noch nicht abschließend geklärt.

Insbesondere ist die gemeinsame Initiative bzw. Arbeitsgruppe zur Abstimmung von Synergieeffekten bezüglich der verschiedenen LIMS und Biomaterialkataloge zu erwähnen. Damit sollen DZG-übergreifend Abstimmungen zur Weiterentwicklung und Zusammenarbeit stattfinden. Die Wahl einer geeigneten Biomaterialverwaltung soll individuell in den einzelnen DZGs erfolgen.

Auch wenn nun viele Standorte den Wechsel auf ein kommerzielles Produkt vollzogen haben, so 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 wenigen Standorten etabliert. Auch scheint die Datenschutzfrage bei vielen Universitätsklinik-Biobanken an der Schnittstelle zwischen Versorgung und Forschung noch ungeklärt zu sein.

### **International**

Im Rahmen der 2012 durchgeführten Umfrage der ISBER WG Informatics [25] wurden 53 Fragen gestellt. Von den 71 Teilnehmern der Umfrage gaben 46% an, von Universitätskliniken zu stammen, 32% aus dem Krankenhausumfeld und 26% aus Forschungslaboren (eine Mehrfachnennung war hier möglich). Von den Befragten werden 44% durch eine separate IT-Abteilung innerhalb der eigenen Organisation unterstützt bzw. 25% von einem IT-Partner außerhalb der Organisation. Bei knapp einem Viertel der Befragten gibt es spezielles IT-Personal in der Biobank selbst. Insgesamt gesehen, ist die IT-Unterstützung von Biobanken also als gut einzustufen.

Bei der Frage nach dem Management der Biobank entfielen 92% der Antworten auf ein Computer-basiertes Datenbankmanagementsystem und von je einem guten Drittel wurde noch „electronic spreadsheet“ bzw. Papier-basiert als Antwort genannt. Auch hier waren Mehrfachantworten erlaubt. Aus diesen Angaben geht leider nicht klar hervor, welche Teile der jeweiligen Biobank schon durch ein Datenbanksystem und welche noch Papier-basiert verwaltet werden. Die Zufriedenheit mit dem jeweiligen aktuell eingesetzten Informationssystem variierte dabei deutlich zwischen sehr zufrieden (15%), einigermaßen zufrieden (41%), weder zufrieden noch unzufrieden (6%), einigermaßen unzufrieden (28%) und sehr unzufrieden (10%).

Insgesamt setzen 48% der Befragten auf eine Eigenentwicklung, rund 30% auf kommerzielle Produkte und die restlichen 22% auf Open-Source-Software. Von den Umfrageteilnehmern, die auf kommerzielle Produkte gesetzt haben, betrug die initialen Kosten zur Beschaffung inklusive Anpassungen und Installation zu je einem Drittel zwischen 10.000 und 50.000 US\$, zwischen 50.000 und 100.000 US\$ und über 100.000 US\$. Die jährlichen Kosten für Wartung und Support liegen bei 50% zwischen 5.000 und 20.000 US\$, bei 25% über 50.000 US\$ und der Rest dazwischen bzw. darunter. Die Implementierung eines kommerziellen Produktes inklusive Validierung und Training dauerte bei mehr als 50% der Befragten 12 Monate und mehr, beim Rest darunter.

Bei Softwareprodukten zur Biomaterialverwaltung wird eine Schnittstelle bzw. Datenintegration mit Datenerfassungssystemen für klinische Informationen der Biomaterialspender als besonders wichtig angesehen. Die Anbindung an Laborgeräte halten zwar 68% für wichtig, aber nur 29% haben sie nach eigenen Angaben bereits realisiert.

Zu den Funktionen, die bei einer Biomaterialverwaltungssoftware als sehr wichtig angesehen werden, gehören:

- ▶ Zuweisung einer Lagerortes (88%)
- ▶ Nachverfolgung der Probenverarbeitung (82%)
- ▶ Plausibilitätschecks bei der Dateneingabe (70%)
- ▶ Barcodes ausdrucken und scannen (69%)
- ▶ Datenschutzbestimmungen einhalten (67%)
- ▶ Versenden und Empfangen von Proben (63%)
- ▶ Konfigurierbarkeit (60%)

Als weniger wichtig betrachtet wurden die Integration von RFIDs sowie das Vorhandensein von APIs zur Systemintegration. Trotz der formulierten Anforderungen war das Ergebnis einer weiteren Frage, dass viele der verwendeten Systeme entweder noch nicht oder nicht zur Zufriedenheit der Anwender in der Lage sind, die oben genannten Funktionen abzubilden.

Bei diesen Ergebnissen sollte beachtet werden, dass ausschließlich ISBER-Mitglieder an der Umfrage teilgenommen haben. Diese sind von sich aus interessiert an dem Thema Biobanken und verfügen deswegen vielleicht auch über eine überdurchschnittliche IT-Unterstützung. Eine Angabe über die jeweilige Nationalität oder den Herkunftsort der Befragten gab es leider nicht [25].

## Bewertung und Handlungsbedarf

Bei zwei der prominentesten Förderlinien von Biobanken, den zentralisierten Biobanken und den CCC-Biobanken, ist die Entscheidung bezüglich eines Softwareproduktes zur Verwaltung des Biomaterials zum größten Teil gefallen. Die nationalen Favoriten scheinen hier CentraXX Biobank und Starlims zu sein. Starlims-unterstützte Biobanken sind meistens schon produktiv, während durch CentraXX-unterstützte Biobanken sich meist noch in der Customizing-Phase befinden. Dies hängt allerdings auch mit den unterschiedlichen Beschaffungszeitpunkten der Software zusammen.

Bei den Deutschen Zentren der Gesundheitsforschung ist die Situation unklarer als bei den oben genannten Biobankvertretern. Aufgrund der besonderen Struktur einiger dieser Verbundforschungsprojekte werden die Biomaterialien teilweise vor Ort gelagert und verwaltet. Zentral findet nur eine Zusammenführung von Metadaten statt. Im Gegensatz zu einer zentralen Erfassung ist in diesen Fällen keine Anschaffung einer zentralen Biomaterialverwaltungssoftware zu erwarten.

Insgesamt gesehen ist der Markt an kommerziellen Produkten immer noch in Bewegung. Die Ende 2009 im Rahmen des TMF IT-Strategie Projektes in Teilprojekt 3 durchgeführte Marktanalyse hatte 21 Softwareprodukte identifiziert. Nach erneuter Prüfung, ob es diese Produkte noch gibt, blieben Mitte 2014 lediglich elf Produkte übrig. Eine zusätzliche Recherche führte zu insgesamt 28 kommerziell verfügbaren Produkten auf dem weltweiten Markt [26]. Unter diesen waren aber lediglich fünf Produkte (CentraXX Biobank, Starlims, BioARCHIVE, GENOMatch und eBioControl) mit Supportstandorten in Deutschland.

Wie auch die internationale Befragung zeigt [25], waren 2012 immerhin 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.

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 hat sich 2014 zum zweiten Mal getroffen.

In Bezug auf viele andere Biobankstandorte scheint es aber immer noch abgekapselte Aktivitäten zur Einführung kommerzieller Softwareprodukte und deren Parametrierung im Hinblick auf die optimale Unterstützung der jeweils lokalen Arbeitsprozesse zu geben. Obwohl konkrete Umsetzungen jeweils in direktem Austausch mit den jeweiligen Softwareentwicklern diskutiert werden müssen, sehen wir doch den Bedarf für einen stärkeren IT-System-unabhängigen Erfahrungsaustausch der deutschen Biobanken.

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 sehr frühen Phase. Erst wenn diese Prozesse umfassend analysiert und definiert wurden, ist eine optimale IT-Unterstützung aller Teilprozesse des Biobanking durch Softwarelösungen möglich.

Darüber hinaus werden Biobankmanagementsysteme an vielen Standorten zunächst 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 Probandaten derzeit noch nicht überall konsequent umgesetzt zu sein. Bei der bereits erwähnten Umfrage zur Biomaterialverwaltungssoftware im Oktober 2014 in der AG Biobanken 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-Tools, z. B. der TMF<sup>45</sup> anbieten. Je nach Umfeld der Biobank, ist dies aber ggf. an der Quelle gar nicht notwendig. Befindet sich die Biobank in einem Versorgungskontext und das Personal hat sich zur Geheimhaltung verpflichtet, so müssen die Proben erst bei der Herausgabe an Dritte pseudonymisiert werden. Die effiziente Umsetzung der hohen Datenschutzanforderungen, wie z. B. die in bestimmten Fällen erforderliche zusätzliche Umschlüsselung der Proben-Identifikation in eine LabIDtr gemäß TMF-Datenschutzkonzept [4], stellt viele Biobanken ebenfalls noch vor große Herausforderungen.

---

45 Link zum Download auf der TMF-Webseite: <http://www.tmf-ev.de/Produkte/P015011>

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). 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 halten wir in der noch frühen Phase der Etablierung von Biobankmanagementsystemen 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>46</sup> und soll in ähnlicher Form auch in 2015 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 ist, die keinem TMF-Mitgliedsverbund angehören.<sup>47</sup>

## 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., Skrownny, 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, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
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.

---

46 vgl. <http://www.pg-ss.imi.uni-erlangen.de/Seiten/11Sitzung.aspx>

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

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. Kiehntopf, M., Böer, K., *Biomaterialbanken – Checkliste zur Qualitätssicherung*. 2008, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
9. Pommerening, K., *Das Datenschutzkonzept der TMF für Biomaterialbanken (The TMF Data Protection Scheme for Biobanks)*. *it - Information Technology*, 2007. **49**(6): S. 352–359.
10. Simon, J.W., Paslack, R., Robiński, J., Goebel, J.W., Krawczak, M., *Biomaterialbanken – Rechtliche Rahmenbedingungen*. 2006, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
11. Hewitt, R., Watson, P., *Defining biobank*. *Biopreserv Biobank*, 2013. **11**(5): S. 309–315.
12. 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.
13. Becker, R., Ihle, P., Pommerening, K., Harnischmacher, U. *Ein generisches Datenschutzkonzept für Biomaterialbanken (Version 1.0)*. 2006, <http://www.tmf-ev.de/produkte/P010021> (Abruf: 2014-06-10).
14. Ethikrat *Humanbiobanken für die Forschung*. 2010. Deutscher Ethikrat, <http://www.ethikrat.org/dateien/pdf/stellungnahme-humanbiobanken-fuer-die-forschung.pdf> (Abruf: 2014-06-11).
15. 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.
16. NCI *NCI Best Practices for Biospecimen Resources*. 2011. National Cancer Institute, <http://biospecimens.cancer.gov/bestpractices/2011-NCIBestPractices.pdf> (Abruf: 2013-04-19).
17. 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: 2013-04-19).
18. 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: 2013-04-19).

19. 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: 2014-01-10).
20. WMA *Deklaration von Helsinki*. 2008. World Medical Association, <http://www.bundesaerztekammer.de/downloads/DeklHelsinki2008.pdf> (Abruf: 2014-06-11).
21. 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: 2013-04-23).
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.

# Molekularbiologische Daten aus Hochdurchsatz-Analysen

*Nadine Umbach, Tim Beißbarth, Ulrich Sax*

## 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 Disziplinen 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.

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 weniger Tage für unter 1000 US\$ entschlüsseln können. Neben dem **Whole-Genome Sequencing (WGS)**, d. h. der Entschlüsselung des 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 gleichzeitige Sequenzierung aller für eine bestimmte Krankheit relevanten Gene, weite Verbreitung. Zur Genotypisierung von Panels genetischer Varianten sind auch die sogenannten **Single Nucleotide Polymorphism (SNP)-Chips**<sup>48</sup> verbreitet im Einsatz, welche bis zu mehrere Millionen

---

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

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.

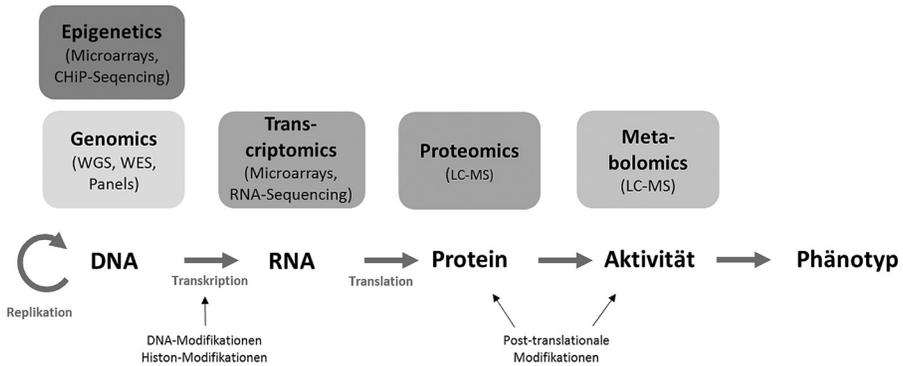


Abbildung 7: Darstellung des zentralen Dogmas der Biologie mit der methodischen Einbindung der Omics-Disziplinen: auf der Hauptachse ist stark vereinfacht der biologische Verlauf von der Replikation der DNA über die Transkription in RNA, die Translation in Proteine bis hin zum klinischen Phänotypen verzeichnet. 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).

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;
- ▶ Metagenomanalyse bei der Untersuchung von Genomfragmenten und deren Zusammensetzung in einer Organismenpopulation und deren Charakterisierung in einer Bioprobe.

Einsatzgebiete in der anwendungsorientierten Forschung finden sich in der 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 ergiebigen 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, aber auch für die Steuerung von Therapieentscheidungen und die Abschätzung von Prognosen [2].

Trotz der Verbesserungen bei der 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 Nationalen Kohorte 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. 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 liegen, 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. 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 beim Einsatz von Omics-Technologien für Forschungszwecke müssen 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 Dienstleistungen (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 bzw. festen 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. Entspricht 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 2 zu entnehmen.

**Tabelle 2: Datenformate und –mengen bei WGS-Analysen**

Work flow	Daten-format	Rechen-zeit	Daten-menge	Kommentar
Bild-daten vom Scanner	Images (TIFF)	bisher 1-10d/run, seit kurzem 1-4d/run	bis zu 300 GB/run	Herstellerspezifische Bilddaten, die <i>on-the-fly</i> erstellt und analysiert werden, und deren Informationen in BCL-Dateien gespeichert werden. Nach der Erzeugung der BCL-Daten werden die TIFF-Dateien standardmäßig gelöscht.

– Tabelle 2 fortgeführt –

– Tabelle 2 fortgeführt –

Work flow	Daten-format	Rechen-zeit	Daten-menge	Kommentar
gefundene Basen	Base Calls Files (*.bcl)	1-10d/run	1,2-1,4 MB pro gezippter BCL-Datei	Herstellerspezifisch: Bei Illumina BCL Binärformat zur Repräsentation der Base Calls. Beinhaltet <i>base calls</i> und Qualitätsparameter für jeden Lauf. <sup>49</sup>
Rohsequenzdaten mit Qualitätsparametern	FASTQ.gz	½-1d/run	bisher bis zu 700 GB/run; neuerdings mit v4-Chemie Verdopplung	Komprimiertes (Gzip-) textbasiertes Format zur Speicherung sowohl von Nukleotidsequenzen als auch der korrespondierenden Qualitätsscores. Dabei wird sowohl das Nukleotid als auch der Qualitäts-Score ( <i>Phred-like score</i> ) 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 CASAVA <sup>50</sup> erfolgen.
Mapping auf Referenzsequenz	SAM (Sequence Alignment / Map) format	30min pro Exom bis 10h pro Genom	bisher bis zu 2,8 TB/run, neuerdings Verdopplung	Generisches textbasiertes Format für die Dokumentation spezifischer <i>Alignments</i> (Übereinstimmungen) verschiedener Sequenzierungsplattformen im Vergleich mit einer Referenzsequenz [11].

– Tabelle 2 fortgeführt –

– Tabelle 2 fortgeführt –

Work flow	Daten-format	Rechen-zeit	Daten-menge	Kommentar
kompri-mierung der Daten	BAM Binary Align-ment / Map)	7h-10h pro Genom	Bisher bis zu 450 GB/run, neuerdings Ver- doppelung	Komprimierte binäre Version von SAM.
Verzeich-nen der Varianten	Variant Call For- mat (VCF)	ca. 100h für Zielregion- Joint-Variant calling (64 Exo- me, 1 Run). 11h-360h pro Genom je nach Varianten und Software	ca. 425 MB (un- komprimiert) für eine Joint Vari- ant Calling-Datei über 64 Exome (1 Run). ca. 1 GB (unkomprimiert) pro Genom	Generisches textbasier- tes Format zur Speiche- rung von DNA-Varianten (SNVs, <i>Indels</i> , Struktur- varianten). Üblicherwei- se komprimiert [12].
	Genome Variation Format (GVF)	Länger als für VCF-Datei (un- terschiedlich: je nach Zielregi- onen und Opti- onen)	Grösser als für VCF-Datei (un- ter-schiedlich: je nach Zielregi- onen und Opti- onen)	Anstelle von VCF-Datei. De facto Standard für die Beschreibung von An- notationsdaten, Mapping zu Ontologien [13].

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 der gesetzlichen Fristen von mindestens 10 Jahren<sup>49, 50</sup> 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 ent-

49 <http://www.illumina.com/documents/products/technotes/RNASeqAnalysisTopHat.pdf>

50 [http://support.illumina.com/sequencing/sequencing\\_software/casava.ilmn](http://support.illumina.com/sequencing/sequencing_software/casava.ilmn)

sprechende 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.

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 und standardisiert erhoben werden. Nur dann können Daten aus verschiedenen Quellen oder von verschiedenen Erhebungszeitpunkten in Relation gesetzt oder für weitere Analysen als gemeinsame Datenbasis verwendet werden [14, 15]. 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) des *Genomic Standards Consortium* [16] und dem *Sequence Read Archive* (SRA) Schema des *European Nucleotide Archive* wurden zwar erste standardisierte Beschreibungen für DNA-Sequenzen entwickelt. In der Praxis finden beide Annotationsschemata jedoch wenig Anwendung. Weiterhin sind die *Minimum Information About a Microarray Experiment* (MIAME) für Microarray-Daten [17], die *Minimum Information About a Proteomics Experiment* (MIAPE) [18, 19], 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 [20, 21], was hinsichtlich der Interpretierbarkeit der Ergebnisse ein ernstes Problem darstellt.

Während Labor-Informations-Management-Systeme (LIMS) im Biobanken-Sektor in den letzten Jahren zunehmende Verbreitung gefunden haben<sup>51</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 von derartigen Systemen beobachtet werden.<sup>52</sup>

---

51 siehe auch das Kapitel zu Biobanken im vorliegenden Bericht

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

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

### **Ethische und rechtliche Rahmenbedingungen**

Um das Potential neuer Technologien für die patientennahe Forschung und Versorgung möglichst weitgehend nutzen zu können, müssen die ethischen und rechtlichen Konsequenzen möglichst frühzeitig erkannt und diskutiert werden. Dies wird beispielsweise in dem BMBF-geförderten Projekt „GenoPerspektiv“<sup>53</sup> aus der ELSA-Initiative und im Rahmen des Projekts „EURAT“<sup>54</sup> untersucht. In der Genomforschung stellen sich ethische und rechtliche Fragen in besonderer Form, da sich Informationen über Geschlecht, Herkunft und Krankheitsdispositionen aus dem Genom ableiten lassen. Insbesondere in der Kombination mit phänotypischen bzw. biometrischen Daten lassen sich diese Datensätze praktisch nicht mehr anonymisieren.

Darüber hinaus gelten genetische Daten aufgrund ihrer besonderen Persönlichkeitsrelevanz, ihrer „Schicksalhaftigkeit“ und der zwangsläufigen Mitbetroffenheit genetisch Verwandter als besonders sensibel. Um einer möglichen sozialen Stigmatisierung oder Diskriminierung zu begegnen, unterliegen diese Daten einem besonderen Schutzbedarf.

Zentrale Voraussetzung der Erhebung und Nutzung molekularbiologischer Daten ist, wie in anderen Forschungsbereichen auch, die informierte Einwilligung. 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

---

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

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

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 [22, 23]. Auch fehlt bisher eine systematische Einschätzung zum aktuellen Wissensstand, den moralischen Einstellungen sowie den Erwartungen und Ängsten der Probanden. Dies wird beispielsweise auch in den regelmäßigen ELSA-Ausschreibungen des Bundesministeriums für Bildung und Forschung (BMBF) thematisiert.

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 geregelt ist, 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 [24]. Werden im Rahmen von Forschungsprojekten oder Studien genetische Daten erhoben und analysiert, können grundsätzlich auch gesundheitsrelevante Befunde außerhalb der eigentlichen wissenschaftlichen Fragestellung aufkommen. Herausforderungen ergeben sich dabei, da oft Veranlagungen für Krankheiten diagnostiziert werden, die beim Probanden noch nicht aufgetreten sind, erst seine Nachfahren treffen können oder für die derzeit keine Behandlungsmöglichkeiten existieren. Somit hat jeder Proband das Recht zu entscheiden, ob er über die Analyseergebnisse und damit über mögliche Krankheitsdispositionen informiert werden möchte oder nicht. 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 [25].

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. 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 [26], führen dazu, dass heute auch von einer faktischen Anonymisierbarkeit genetischer Daten kaum noch ausgegangen werden kann.

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 [14]. 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.

## 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 [27];
2. Produkte an der Schnittstelle vom Datenproduzenten zum Auswerter (s. auch Review von Canuel [28]);
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 practices, 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 im Kontext externer Informationsdienste aus der Systembiologie abgefragt und 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.

Ansätze zur datenschutzgerechten Gestaltung von IT-Architekturen für die Integration von Omics-Daten in die verschiedenen Forschungslandschaften werden in [29] 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 [27]):
  - 1.1 Analyse
    - ▶ Bioconductor; Erweiterungen von R (<http://www.bioconductor.org>)
    - ▶ BioDAS; Strukturbiologie (<http://www.biodas.org>)
    - ▶ 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>55</sup>; Portal mit Werkzeugen und Workflow-Management für Systembiologie und Cheminformatik; kann kommerzielle Datenbanken wie TRANSPATH® und TRANSFAC® mit einbinden.
  - ▶ KNIME; Pipeline Tool (<http://www.knime.org/knime>)
  - ▶ R; Analyse (<http://www.r-project.org>)
  - ▶ Taverna; Analysewerkzeug (<http://www.taverna.org.uk>)
- 1.2 Ablage
- ▶ Adama; Datenbank für heterogene Daten (<http://www.systemsbio.org/addama>)
  - ▶ BioMart; Datenbanksystem für biologische Daten (<http://www.biomart.org>)
  - ▶ openBIS [30]; offenes verteiltes System für das Management biologischer Daten ([www.cisd.ethz.ch/software/openBIS](http://www.cisd.ethz.ch/software/openBIS))
  - ▶ SDCubes; Datenbank für heterogene Daten (<http://www.semanticbiology.com/software/sdcube>)
2. Produkte an der Schnittstelle vom Datenproduzenten zum Auswerter (s. auch Review von Canuel [28]):
- ▶ openBIS [30]; offenes verteiltes System für das Management biologischer Daten ([www.cisd.ethz.ch/software/openBIS](http://www.cisd.ethz.ch/software/openBIS))
  - ▶ SEEK [31]; Plattform zum Austausch von Daten und Modellen in der Systembiologie, soll von FAIRDOM mit openBIS in openSEEK zusammengeführt werden.
  - ▶ tranSMART [32]; ein zunächst aus i2b2 [33, 34] 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 [28].
3. Lösungen für Omics-Daten mit Wissensmanagement-Perspektive:
- ▶ BioXM<sup>TM</sup><sup>56</sup> [35, 36]; Umgebung zur Integration und Analyse biomedizinischer Daten in eine Wissensmanagement-Umgebung.
  - ▶ eTRIKS<sup>57</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.

---

55 <http://genexplain.com/genexplain-platform-1>

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

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

4. Initiativen für besseres Datenmanagement in der biomedizinischen Informatik bzw. Systembiologie:
- ▶ FAIRDOM<sup>58</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>59</sup>; Ziel ist der schnellere Austausch von Forschungsdaten; bisher Schwerpunkt „Data Foundation and Terminology Working Group“.

## 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:

- ▶ 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.

---

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

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

- ▶ 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>60</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.

---

60 S. Empfehlung 7: Sicherung und aufbewahrung von Primärdaten, Denkschrift „Sicherung guter wissenschaftlicher Praxis, Safeguarding Good Scientific Practice“ DFG 1998, 2013

[http://dfg.de/en/research\\_funding/principles\\_dfg\\_funding/good\\_scientific\\_practice/index.html](http://dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html)

- ▶ 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 abgedeckt 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. Konkrete und möglichst generische Konzepte für die Integration von NGS-Analysen in die klinische Routine müssen entwickelt und umgesetzt werden.
- ▶ 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

Für den fachlichen Rat bei der Erhebung der aktuellen IT-Landschaft bei Omics-Technologien danken wir:

- ▶ Jürgen Eils, Heidelberg
- ▶ Michael Forster, Kiel
- ▶ Prof. Dr. Andre Franke, Kiel
- ▶ Dr. Harald Kusch, Göttingen
- ▶ Benjamin Löhnhardt, Göttingen

Diese Arbeit wurde neben der Förderung durch die TMF<sup>61</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.

---

61 Die Erstellung dieses IT-Reports durch die TMF wurde wiederum vom BMBF im Rahmen des Projekts MethInfra-Net (FKZ 01GI1003) gefördert, s. auch einleitendes Kapitel.

## Literatur

1. Hg, 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.
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: 2015-02-20).
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. Reese, M.G., Moore, B., Batchelor, C. et al., *A standard variation file format for human genome sequences*. Genome Biol, 2010. **11**(8): S. R88.
14. DFG Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, Unterausschuss für Informationsmanagement: *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*. 2009. Deutsche Forschungsgemeinschaft, [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: 2015-02-20).

15. 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.
16. 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.
17. 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.
18. Taylor, C.F., *Minimum reporting requirements for proteomics: a MIAPE primer*. Proteomics, 2006. **6 Suppl 2**: S. 39–44.
19. 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.
20. McDermott, J.E., *Reproducibility: two more red flags for suspect work*. Nature, 2013. **499**(7458): S. 284.
21. Begley, C.G., *Six red flags for suspect work*. Nature, 2013. **497**(7450): S. 433–434.
22. 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.
23. Wolf, S.M., *The past, present, and future of the debate over return of research results and incidental findings*. Genet Med, 2012. **14**(4): S. 355–357.
24. Duttge, G., *Das Recht auf Nichtwissen in der Medizin*. Datenschutz und Datensicherheit – DuD, 2010. **34**(1): S. 34–38.
25. NER *Biobanken für die Forschung*. 2004. Nationaler Ethikrat, [http://www.ethikrat.org/dateien/pdf/NER\\_Stellungnahme\\_Biobanken.pdf](http://www.ethikrat.org/dateien/pdf/NER_Stellungnahme_Biobanken.pdf) (Abruf: 2014-06-11).
26. Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y., *Identifying Personal Genomes by Surname Inference*. Science, 2013. **339**(6117): S. 321–324.
27. Berger, B., Peng, J., Singh, M., *Computational solutions for omics data*. Nat Rev Genet, 2013. **14**(5): S. 333–346.
28. 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.
29. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten – Generische Lösungen der TMF 2.0*. 2014, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.

30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.



# Mobile IT-Werkzeuge

*Karoline Buckow, Otto Rienhoff*

## Anwendungsbereich und Eingrenzung

### **Mobile IT-Werkzeuge im Gesundheitswesen**

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. Der seitdem erfolgte Ausbau der Mobilfunknetze, die Entwicklung besserer Techniken und Standards für die Datenübertragung sowie immer leistungsfähigerer Sensoren und Telekommunikationsgeräte erlaubt es, zunehmend mobile IT-Werkzeuge in der Versorgung einzusetzen, etwa zu Verlaufsbeobachtungen. Dabei entstehen zum Teil grundsätzlich neue Methoden, mit denen Prozesse in der Versorgung quantifiziert werden können. Als Beispiel sei hier die chemische und physikalische Vermessung chronischer Wunden genannt oder das breite Spektrum von Lab-on-a-Chip-Anwendungen, die sich zunehmend in der ambulanten Versorgung und der medizinischen Forschung zu etablieren beginnen [2]. 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, 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. 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 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 97.000 Apps von den Herstellern der Kategorie mHealth zugeordnet [3]. 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 [4], 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

## **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. 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, um ggf. beschreibende Metadaten zu dokumentieren. 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 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 [5, 6].

Mit einer solchen Erfassung sogenannter Patient Reported Outcomes (PRO) sollen Probleme von Patienten bezüglich des Befindens oder des Alltags früh erkannt, Schlussfolgerungen auf erforderliche therapeutische Maßnahmen ermöglicht und eine Verbesserung der Lebensqualität und Zufriedenheit mit der Gesundheitsversorgung erreicht werden [7]. Mobile Geräte wie Smartphones oder Tablets und die entsprechenden Applikationen können aufgrund ihrer zunehmenden Verbreitung ein geeignetes Werkzeug für die patientenorientierte Datenerfassung sein.

### **Einsatz mobiler Sensoren**

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 [8]. Und der Markt für Devices und Dienste, die mobile Sensoren einsetzen und verarbeiten, steigt – mit erheblichen Auswirkungen auf die damit verbundene Datenmenge [3, 8]. 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 [3].

### **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. 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 [3]. Diese Situation führt zu Konsequenzen hinsichtlich der Nachhaltigkeit, Professionalität und Interoperabilität bestehender Entwicklungen.

## 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) 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 [9], den FDA Code of Federal Regulations 21 CFR Part 11 [10] sowie die FDA Guidance for Industry - Computerized Systems Used in Clinical Investigations [11] 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 [12]. 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:

1. 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.
2. Anforderungen an die Kommunikation / Interoperabilität: Für eine Kommunikation der App mit weiteren Forschungsinfrastrukturkomponenten müssen Softwareseitig 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.

3. 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.
4. 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.
5. 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.
6. Anforderungen an die Metadatendokumentation: die neuen Möglichkeiten bergen auch neue Gefahren für Fehlerhebungen oder geräte- bzw. prozessabhängige Fehler. Deshalb müssen die Verfahren diesen Qualitätsaspekt nachvollziehbar adressieren.

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

Sowohl das im Rahmen des Forschungsprojekts 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.

Je nach Einsatz der App und in Abhängigkeit vom gewählten Einsatzbereich müssen unter Umständen das Arzneimittelgesetz oder das Medizinproduktegesetz bei der Konzeption der App berücksichtigt werden. Die Europäische Kommission hat in einer im Januar 2012 veröffentlichten Guideline Entscheidungskriterien definiert, die Applikationen für Medizinprodukte (engl. medical devices) entsprechend geltender Richtlinien [13, 14] gegenüber Nicht-MPG-Applikationen abgrenzen [15].

## Kategorisierung der IT-Landschaft

Bei der Einteilung der IT-Landschaft werden im Folgenden 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 von mobilen Devices abgelöst werden. Die dritte Kategorie umfasst mobile IT-Werkzeuge, die am oder im Körper getragen werden. Hierzu zählen unter anderem Smartwatches oder Implantate, die Messwerte über Sensoren aufzeichnen und versenden.

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. Bei der Planung einer Studie und der Datenerfassung über eine Studiensoftware werden Visitenpläne vorgegeben, die fixe Beobachtungszeitpunkte vorgeben. 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 den zugehörigen Arzt-dokumentierten Daten in der Studiendatenbank setzt ein übergreifendes Identitätsmanagement voraus. Entsprechende Schnittstellen und Prozesse, die für die Kommunikation zwischen Studiensoftware und ei-

nem 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 ist entsprechend ausschlaggebend für die Notwendigkeit von Geräteschnittstellen. Ist im Rahmen des Forschungsprojekts neben der Datenerfassung durch den Patienten eine Aufnahme von Daten aus Messinstrumenten oder von Sensoren erforderlich, müssen entsprechende Geräteschnittstellen implementiert, getestet und validiert werden.

## Aktueller Stand

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

Im Jahr 2009 veröffentlichte die Arzneimittelzulassungsbehörde der USA, die Food and Drug Administration (FDA), eine Guidance for Industry zum Thema Patient Reported Outcome (PRO). Darin fordert die FDA eine stärkere Verwendung von PROs, um

1. unterschiedliche Wahrnehmungen gesundheitsrelevanter Aspekte von Arzt und Patienten erkennen zu können,
2. durch den Erinnerungs-Bias verschwommene Schilderungen zu vermeiden und
3. die Reliabilität von Informationen durch Standardisierung (im Vergleich zu Arztgeführten Interviews) zu optimieren [5].

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 [16]. 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 CFR Part 11 [10] auch für die elektronische Erfassung von PRO-Daten über mobile Geräte hin [16].

Im April 2014 veröffentlichte die Europäische Kommission ein Grünbuch zum Thema mobile Health-Dienste [3]. 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 Leistungsnachweise, die Maßnahmen zur Prüfung von mHealth-Diensten und die Optimierung der Interoperabilität von mHealth-Diensten [3].

## **Erfassung von PRO-Daten**

Aktuelle Publikationen zeigen die Vorteile einer PRO-Datenerfassung über elektronische Medien gegenüber einer papierbasierten Datenerfassung auf. Hierzu zählen die Reduzierung fehlender Angaben [17], die Verhinderung invalider oder unvollständiger Eingaben durch die Möglichkeit der gesteuerten Datenerfassung (z. B. über Sprungmarken) [16, 17], die Eingrenzung der Antwortoptionen (z. B. über Radio-Buttons, durch die nur eine Antwortoption zugelassen wird) zur Vermeidung mehrdeutiger Antworten [17], die Steigerung der Compliance durch die Protokollierung sämtlicher Aktionen in einem Audit Trail [17, 18] sowie die Optimierung des Datenmanagements und der Datenqualität [19]. Tablet-PCs vereinen die Vorteile papierbasierter Dokumentation in Bezug auf die Einfachheit und elektronischer Datenerfassung in Bezug auf das Datenmanagement und die Datenqualität, so das Ergebnis einer vergleichenden Studie über verschiedene Methoden der PRO-Datenerfassung [20]. Sie punkten vor allem in Bezug auf die Einfachheit der Anwendung, die Flexibilität und die Genauigkeit / Fehlerrate gegenüber anderen Methoden [20]. 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 [21].

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 bietet hierfür eine technische Grundlage – die Eignung von Apps für dieses Anwendungsfeld muss 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 [22].

## **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 [23]. 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 [23]. 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 [9]. 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 [23]. Grundsätzlich gilt, dass die Nutzung von Daten aus mHealth-Diensten für die Forschung konform mit geltenden Regularien erfolgen muss. Für Forschungsvorhaben auf personenbezogenen Daten, die ursprünglich zu einem anderen Ziel erfasst wurden, ist eine entsprechende fehlende Einwilligung durch den Patienten problematisch [3]. 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 [24]. Darüber hinaus mangelt es bislang an standardisierten Lösungen für den Nachweis über die Funktionstüchtigkeit der App, die Korrektheit der Funktionsweise bzw. über die Durchführung einer angemessenen Testung [3].

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 [25]. Während der Übergang vom Dienstleistungs- zum Sekundärnutzungsbereich weitgehend gemäß der technischen und organisatorischen Vorgaben der generischen Datenschutzkonzepte der TMF [26] 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 [25].

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>62</sup> Nach wie vor sind die Gesetze und Regularien nicht an die Bedürfnisse des Einsatzes von Medical Devices im mobilen Sektor (mHealth, Telemedizin, AAL) angepasst und verlangsamten dadurch das effektive Voranschreiten entsprechender Entwicklungen. Eine Chance, Standards im Datenschutz für mobile Gesundheitsanwendungen zu schaffen, könnte das im Entwurf vorliegende eHealth-Gesetz darstellen. Es gilt jedoch noch kritisch zu reflektieren, inwiefern diese Gesetzesinitiative die gewünschte Entwicklung unterstützen kann.

### **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 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 [16]. Die FDA-Guidance für den Umgang mit PRO-Daten definiert, dass dies Überlegungen zur statistischen Auswertung von PRO-Daten einschließt [16]. 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 [16]. Apps bieten hier die Möglichkeit, den Probanden zu definierten Zeitpunkten an die Daten-

---

62 [www.tmf-ev.de/news/1690](http://www.tmf-ev.de/news/1690)

erfassung zu erinnern oder auf fehlende Angaben hinzuweisen, um der Zahl fehlender Werte entgegenzuwirken.

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 [27].

Eine Prüfung des Einsatzes von Apps muss für jede Studie separat erfolgen. Akzeptanz, räumliche 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 sind ein vielversprechendes Werkzeug für die patientenorientierte Datenerfassung. Sie vereinen die Vorteile der computergestützten Datenerfassung mit einer intuitiven Nutzeroberfläche und nahezu ständiger Verfügbarkeit. Ein professioneller Einsatz im Rahmen wissenschaftlicher Forschung erfordert jedoch eine eingehende Evaluation des Ansatzes. Die Interpretation der erfassten Daten erfordert ein Verständnis über den Umgang des Patienten mit dieser Technologie im Vergleich zu anderen Erfassungsmethoden. Zudem kann eben dieser Umgang 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.

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, zum einen hinsichtlich technisch innovativer Lösungen zur Abschottung und Absicherung von Apps gegenüber der Betriebssysteminfrastruktur, zum anderen aber auch in Bezug auf eine sorgfältige Abwägung der enormen, ggf. auch für die individuell betroffenen Patienten relevanten, Chancen einerseits mit den genannten Risiken andererseits.

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 zu 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. Schmidt, S., Rienhoff, O., *Interdisciplinary assessment of personal health monitoring*. 2013, Amsterdam: IOS Press.
3. EC *Grünbuch über Mobile-Health-Dienste („mHealth“)*. 2014. Europäische Kommission, <http://ec.europa.eu/digital-agenda/en/news/green-paper-mobile-health-mhealth> (Abruf: 2014-08-27).
4. Heinze, O. *Entwicklung einer iOS App zur Erfassung eines standardisierten elektronischen Notfalleinsatzprotokolls*. 2012, <http://www.telemed-berlin.de/telemed2012/programm-2012-download> (Abruf: 2013-04-19).
5. Ebner-Priemer, U.W., Bohus, M., *Elektronische Tagebücher in der Medizin: Echtzeiterfassung von Symptomen*. Deutsches Ärzteblatt, 2008. **105**(28-29): S. 1550–1552.

6. Wilhelm, P., Perrez, M., *Ambulantes Assessment in der klinischen Psychologie und Psychiatrie*. Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 2008. **56**(3): S. 169–179.
7. Chang, C.-H., *Patient-reported outcomes measurement and management with innovative methodologies and technologies*. Quality of Life Research, 2007. **16**(1): S. 157–166.
8. 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.
9. 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: 2013-04-18).
10. FDA CFR – Code of Federal Regulations Title 21. 1997, <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=11&showFR=1> (Abruf: 2014-08-04).
11. 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: 2013-04-19).
12. Buckow, K. *Anforderungen an eine GCP-konforme Dokumentation für Studien und Forschungsprojekte mit Hilfe von Studien-Apps*. 2012, [http://www.telemed-berlin.de/telemed/2012/vortrag/vortrag\\_buckow454\\_476.pdf](http://www.telemed-berlin.de/telemed/2012/vortrag/vortrag_buckow454_476.pdf) (Abruf: 2013-04-19).
13. 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: 2015-02-21).
14. EP/COE Directive 98/79/EC of the European Parliament and of the Council of 27 October 1998 on in vitro diagnostic medical devices. 2011. European Parliament / Council of Europe, <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:01998L0079-20120111&qid=1413308118275&from=EN>.
15. 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. 2012. European Commission, [http://ec.europa.eu/health/medical-devices/files/meddev/2\\_1\\_6\\_ol\\_en.pdf](http://ec.europa.eu/health/medical-devices/files/meddev/2_1_6_ol_en.pdf) (Abruf: 31.10.2013).

16. 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: 2013-04-19).
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. Mocomed Arbeitskreis *Mobiles Computing in der Medizin*. <http://fg-mocomed.gi.de/> (Abruf: 2013-04-19).
23. Fahrenberg, J., *Ambulantes Assessment, in Handbuch Statistik, Methoden und Evaluation*, Hrsg.: H. Holling and B. Schmitz. 2010, Hogrefe: Göttingen. S. 201–212.
24. 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.
25. Pommerening, K., *Personalisierte Medizin und Informationstechnik - Aspekte des Datenschutzes, in 55. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS)*, 2010: Mannheim, <http://www.gmds2010.de/wp-content/uploads/fohlen/91.pdf>.
26. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
27. Fahrenberg, J., Leonhart, R., Foerster, F., Hrsg. *Alltagsnahe Psychologie. Datenerhebung im Feld mit hand-held PC und physiologischem Mess-System*. 2011, e-Buch.

# Sekundärnutzung klinischer Daten

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

## Anwendungsbereich

Eine Reihe von Projekten aus dem Umfeld der TMF oder auch mit direkter Beteiligung der TMF 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. Zu diesen Projekten gehört die Berliner Forschungsplattform Gesundheit,<sup>63</sup> die Projekte cloud4health<sup>64</sup> [1] und EHR4CR<sup>65</sup> [2] sowie ein BMBF-finanziertes Projekt zur KIS-basierten Unterstützung der Rekrutierung in klinischen Studien<sup>66</sup> [3]. Parallel zu dieser vermehrten Projektaktivität wurde in den letzten Jahren mit Unterstützung durch die TMF die Arbeitsgruppe „Nutzung von elektronischen Patientenakten für die klinische Forschung“ in der GMDS gegründet und etabliert, die sich ebenfalls explizit dem Thema der Sekundärnutzung widmet.

Auch wenn man über den „Tellerrand“ der TMF hinausschaut, kann man feststellen, dass die Sekundärnutzung klinischer Behandlungsdaten enorm an Bedeutung gewonnen hat [4]. Relevante Anwendungsfälle reichen dabei 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 [vergl. 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] und von manchem vielleicht sogar als

---

63 [http://medinfo.charite.de/forschung/berliner\\_forschungsplattform\\_gesundheit/](http://medinfo.charite.de/forschung/berliner_forschungsplattform_gesundheit/)

64 <http://www.cloud4health.de>

65 <http://www.ehr4cr.eu>

66 <http://www.tmf-ev.de/kisrek>

der typischere Vertreter angesehen werden mag, 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 Datentransparenzverordnung mit ihren Öffnungsklauseln für Forschungseinrichtungen und die hierzu von der TMF im Rahmen des Forums Versorgungsforschung durchgeführten Veranstaltungen<sup>67, 68</sup>, zunehmend auch von der akademischen Forschung in den Blick genommen werden, sollte dieser Bereich auch in einem der kommenden Berichte mit untersucht werden.

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, wieviel Patienten mit den relevanten Ein- und Ausschlusskriterien behandelt wurden. Diese Zahlen dienen als bestmögliche Schätzung für künftig ansprechbare 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. Die klinischen Daten werden dann direkt in einer Form und Qualität erhoben und elektronisch

---

67 siehe <http://www.tmf-ev.de/news/1464>

68 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)

erfasst, dass sie sowohl als Dokumentation des Versorgungsprozesses als auch direkt für klinische Studien mitgenutzt werden können [7, 10]. Ein alternativer Workflow sieht eine Vorfürfüllung der elektronischen Studienformulare (eCRFs) durch die in der Behandlung dokumentierten und elektronisch erfassten Daten vor [11, 12]. 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, welche diese Quelldaten liefern, sicherzustellen.

Ein weiteres Anwendungsfeld für die Sekundärnutzung klinischer Daten entsteht durch den systematischen Aufbau von Biobanken an den klinischen Standorten in Deutschland.<sup>69</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 „Individualisierten Medizin“ 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 [13, 14].

## 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.

Die Forschungs-IT wird hingegen immer häufiger zentralisiert und einrichtungsübergreifend betrieben. Im Rahmen kommerzieller Studien sind regelmäßig Pharmafirmen 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

---

<sup>69</sup> vergleiche hierzu auch das Kapitel zu Biobanken im vorliegenden Bericht

– 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>70</sup> 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.

## 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.

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 Betroffener kaum oder gar nicht möglich ist. In diesen Fällen ist dann doch eine ausführliche Betrachtung der gesetzlichen Rahmenbedingungen erforderlich. Diese sind leider unübersichtlicher als oft dargestellt. Zunächst ist

---

<sup>70</sup> vergl. das Kapitel zu klinischen Studien im vorliegenden Bericht

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>71</sup>

Zu Recht wird häufig darauf hingewiesen, dass die Grundzüge des Datenschutzes in den jeweiligen Landesdatenschutzgesetzen und dem BDSG gleichsinnig geregelt sind. Allerdings sind gerade die Forschungsklauseln, die in bestimmten Fällen 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, in den verschiedenen Datenschutzgesetzen 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 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 [15].

---

71 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 [15].

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. Im Herbst 2014 hat der BGH diese grundsätzliche Frage anlässlich eines fraglichen Personenbezugs von IP-Adressen bei Webseitenbetreibern dem EuGH vorgelegt [16].

Das von der TMF beauftragte Gutachten votiert für die Annahme des Konzepts des relativen Personenbezugs, welches besagt, dass Daten in Abhängigkeit von der Verfügbarkeit entsprechenden Zusatzwissens bei einer Stelle als personenbezogen und gleichzeitig bei einer zweiten Stelle als nicht personenbeziehbar gelten können. Entsprechend können pseudonymisierte Daten bei einer Stelle, die selbst nicht die Pseudonyme auflösen kann und auch sonst über keine Möglichkeit der Depseudonymisierung verfügt, als anonymisiert angesehen werden. Diese Sichtweise erlaubt somit die Übermittlung pseudonymisierter klinischer Daten an eine externe Stelle, ohne dass hierfür eine separate gesetzliche Grundlage geschaffen werden müsste. Voraussetzung hierfür ist aber immer, dass die medizinischen Daten selbst nicht so umfangreich sind, dass schon allein anhand dieser Daten eine Reidentifikation einzelner Datensätze möglich ist [15].

In dem Gutachten von Schneider wird diese Frage ausführlich diskutiert und die Antwort vor dem Hintergrund einer umfassenden Literatursichtung detailliert begründet [15]. Der BGH spricht in seiner Vorlage an den EuGH in Bezug auf die im Gutachten vertretene Sichtweise von einer überwiegenden Auffassung [16]. Und doch zeigt die intensive Diskussion der TMF mit den zuständigen Gremien der Datenschützer, dass hierzu bei den relevanten Aufsichtsbehörden in Deutschland nach wie vor unterschiedliche Auffassungen bestehen. Somit kann leider für Forschungsprojekte kein einheitlich für alle Bundesländer geltender Ratschlag hinsichtlich dieser Problematik gegeben werden. Lediglich bei Akzeptanz eines absoluten Personenbezugs ist von einer problemlosen Abstimmung mit allen Aufsichtsbehörden in Deutschland auszugehen. Das bedeutet, dass pseudonymisierte Daten immer als personenbeziehbar anzusehen sind, auch wenn die verarbeitende Stelle selbst keinen Zugriff auf die der Pseudonymisierung zugrundeliegende Zuordnungsregel besitzt.

Gerade das Konzept des absoluten Personenbezugs scheint für viele Szenarien der Sekundärnutzung eine hohe Hürde darzustellen, da in Folge der Anwendung dieses Konzepts in vielen Fällen 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 eingrenzbareren 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. Die gleiche Problematik gilt ebenfalls für die Anonymisierung klinischer Daten, die daher häufig auch eine zusätzliche Vergrößerung der medizinischen Nutzdaten erfordert.

Wenn Daten oder Proben aufgrund einer Einwilligungserklärung der Patienten sekundär für die Forschung genutzt werden sollen, so muss diese Erklärung bestimmt sein, so dass klar zu erkennen ist, unter welchen Bedingungen sich die betroffene Person mit der Erhebung, Verarbeitung oder Nutzung welcher Daten einverstanden erklärt. 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 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].

## 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 auch praktisch unmöglich, zu beurteilen, welche Datenelemente in verschiedenen Datenquellen verfügbar sind und daher für eine gemeinsame wissenschaftliche Auswertung überhaupt in Frage kommen [19].

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 dokumentiert, so dass davon auszugehen ist, dass auch die qualitätssichernden Verfahren dieses Ziel 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; 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 [20].

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 [21]. Inwieweit der bei der WHO in Vorbereitung befindliche internationale Kodierstandard 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 [22, 23]. Insbesondere würde SNOMED-CT eine semantische Annotation von Patientenakten ermöglichen und wäre damit für eine integrative Analyse aus mehreren Datenquellen von großer Bedeutung [24].

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 derzeit sowohl in der Forschung als auch im Kontext der Routineversorgung, einerseits aus lizenzrechtlichen und Kostengründen, andererseits aber auch aufgrund der hohen Komplexität in der Implementation praktisch nicht zum Einsatz [21].

Nun gibt es aber auch weniger umfassende und kostenträchtige sowie einfacher umsetzbare Lösungsansätze, die im Wesentlichen darin bestehen würden, überhaupt vorhandene standardisierte Kodiersysteme vermehrt auch in der Routinedokumentation einzusetzen. Die Hürde für die wissenschaftliche Nachnutzung von Daten ist eben genau dann am höchsten, wenn überhaupt keine standardisierte Kodierung zum Einsatz kommt [25]. So gibt es z. B. mit LOINC für den Bereich der Laboruntersuchungen ein mächtiges terminologisches System, welches zudem lizenzkostenfrei eingesetzt werden kann [26, 27]. Zudem wird in Kombination mit LOINC auch die Verwendung von UCUM für die standardisierte Erfassung der Messgrößen empfohlen [28]. 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 [29]. Auch wenn heute bei Weitem nicht für alle Bereiche der medizinischen Dokumentation passende 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. Gleichwohl ist mittelfristig auch ein mindestens evaluatorischer, stufenweiser Einstieg in die Nutzung von SNOMED-CT zu empfehlen [21].

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>72</sup> untersucht, wieviele der in einer Zufallsauswahl von Studien vorgefundenen Ein- und Ausschlusskriterien tatsächlich auch in der elektronischen

---

72 siehe [www.tmf-ev.de/kisrek](http://www.tmf-ev.de/kisrek)

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 [30].

Auch wenn somit offensichtlich längst nicht alle für die Forschung wichtigen Daten auch schon elektronisch dokumentiert aus dem Behandlungsfall heraus vorliegen und 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 [30]. Es gibt bisher aber leider nur wenige deutsche Projekte, in denen die Erhöhung der Rekrutierungsrate [6] oder die Effizienzsteigerung bei der Rekrutierung bzw. Datenerhebung auch explizit nachgewiesen werden konnte [31].

Neben der Frage der geeigneten Kodierung bzw. eines notwendigen Mappings [32, 33] 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 [34, 35], auch wenn gerade in den akademischen Forschungseinrichtungen noch eine weitergehende Standardisierung wünschenswert wäre [36]. 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 [37]. Mit dem jüngst propagierten modularen, webbasierten Konzept Fast Healthcare Interoperability Resources (FHIR) versucht HL7 aktuell einen neuen Weg in der Standard-Entwicklung einzuschlagen.<sup>73</sup> Es bleibt abzuwarten, welche Auswirkungen dies auf die Landschaft der real genutzten Datenstandards im Bereich der Patientenversorgung in Deutschland haben wird. Seit 2008 arbeiten die beiden Standardisierungsorganisationen HL7 und CDISC im Joint Initiative Council for Standards Harmonization (JIC)<sup>74</sup> zusammen mit der ISO und anderen Institutionen gemeinsam an der Weiterentwicklung der relevanten Standards, so dass auf Basis dieser Kooperation

---

73 siehe <http://hl7.org/fhir>

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

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 derzeit jenseits regulierter klinischer Studien 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 Interoperabilitätsvoraussetzungen 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>75</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 [38]. 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>76</sup>

## 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. Eine komplexe Integration in die Abläufe der Versorgung und damit in bestehende IT-Infrastrukturen der behandelnden Einrichtungen erscheint in diesem Fall nicht nötig. Anders sieht dies hingegen für das Szenario der Rekrutierungsunterstützung in klinischen Forschungsprojekten aus, welches eine komplexe Architektur erfordert und daher im Folgenden beispielhaft für eine mögliche Kategorisierung der IT-Landschaft Modell steht. Ein weiterer Vorteil dieses Vorgehens besteht darin, dass in dem BMBF-geförderten Projekt zur KIS-basierten Rekrutierungsunterstützung eine ausführliche Analyse der beteiligten Systeme und notwendiger Funktionen durchgeführt wurde, anhand derer dann eine generische Architektur entwickelt werden konnte. Die folgenden Ausführungen basieren ganz wesentlich auf diesen Ergebnissen [3, 39, 40].

---

75 [www.ihe.net](http://www.ihe.net)

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

## 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) vorhanden, wenn dieses 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 Spezialsysteme, 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>77</sup> realisieren. Die vielen unterschiedlichen technischen Lösungsvarianten müssen letztlich alle eine standardisierte Abfrageschnittstelle für die Forschungsfragestellungen bieten und dürfen den Routinebetrieb des KIS nicht

---

77 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.

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 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.

### **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“ [41]. 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 [42].

### **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 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, 43]. 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 Universitätsklinika reale Implemen-

tierungen 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 diente beispielsweise an einem Standort das KAS mit einem separaten Tabellenbereich und an einem anderen ein vollständig separates Forschungs-Data-Warehouse [3]. Bei einem dritten Partner übernahm gar der Kommunikationsserver die Rolle der Datenbasis [44].

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. Dieser Abschnitt fokussiert daher den aktuellen Stand von IT-Infrastrukturen in der Versorgung. Allerdings kann hier kein umfassender Überblick über die komplette IT-Ausstattung aller forschungsrelevanten behandelnden Einrichtungen in Deutschland gegeben werden, zumal die Infrastrukturen gerade an den universitären Standorten mit ihren vielen Fachabteilungen und entsprechend spezialisierten Dokumentationssystemen schon für sich zu enormer Komplexität neigen.<sup>78</sup>

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, deren Ergebnisse hier zusammengefasst dargestellt sind. Während für die Universitätskliniken eine Vollerhebung möglich war, basieren die Daten zu den nicht-universitären Krankenhäusern auf einer Stichprobenerhebung [39].

---

78 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 AP1 unter [www.tmf-ev.de/kisrek](http://www.tmf-ev.de/kisrek) dokumentiert ist.

**Tabelle 3: Häufigkeit von Klinischen Arbeitsplatzsystemen an deutschen Krankenhäusern**

Hersteller	Produkt	Uni-Kliniken [N]	Uni-Kliniken [%]	Andere KH* [%]
Siemens AG	i.s.h.med	15	46	14
Agfa-Gevaert Gruppe	ORBIS	8	24	32
Siemens AG	medico	4	12	8
Siemens AG	Soarian	2	6	0
	Eigenentwicklung	2	6	1
	Andere	2	6	36
	Keine Angabe	0	0	9

*Alle Angaben aus dem Bericht zum Arbeitspaket 7 des Projekts “KIS-basierte Unterstützung der Patientenrekrutierung in klinischen Studien” (BMBF-FKZ: 01EZ0941A, [www.tmf-ev.de/kisrek](http://www.tmf-ev.de/kisrek))*

*\* In den anderen Krankenhäusern wurde in der Befragung nicht zwischen KAS und PDV unterschieden. Zudem basieren die Ergebnisse auf einer nicht-repräsentativen Stichprobenerhebung bei 78 Krankenhäusern in der Umgebung der 5 beteiligten universitären Standorte.*

Wie die in Tabelle 3 dargestellten Ergebnisse zeigen, dominieren wenige Software-systeme 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 Siemens<sup>79</sup> und Agfa decken derzeit 88% dieses Markts ab. Im KIS-Rekrutierungsprojekt waren genau diese vier Systeme und an einem fünften Standort noch eine Eigenentwicklung vertreten.

<sup>79</sup> Als Anbieter wird in diesem Bericht noch Siemens gelistet. Im August 2014 wurde Siemens Health Services mit allen Krankenhausinformationssystemen vom amerikanischen Unternehmen Cerner übernommen.

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. Beispielsweise gibt es in zwei der untersuchten Systeme die Möglichkeit, in Übersichtslisten die in einer Abfrage gefundenen Patienten als Studienkandidaten durch ein ausgewähltes Symbol zu markieren. Da aber in den Listen nur Platz für ein Symbol vorgesehen ist, und es auch schon Symbole für andere wichtige Hinweise zu den Patienten gibt, funktioniert eine Benachrichtigung auf Basis dieser generischen Funktion nur eingeschränkt. Zudem fehlen Funktionen zur Abbildung einer Studienverwaltung in allen untersuchten Systemen [39].

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.

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 ermöglichen soll.<sup>80</sup> Als Basis des Werkzeugkastens wurde nach einer Evaluation im Rahmen des IT-Strategie-Projekts der TMF die Plattform i2b2<sup>81</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

---

80 siehe [www.tmv-ev.de/idrt](http://www.tmv-ev.de/idrt)

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

Abfrageschnittstelle verbindet [45]. Seit 2009 hat sich die TMF auch am Aufbau einer zunächst nationalen und mittlerweile europäischen User Group der i2b2-Anwender im akademischen Umfeld beteiligt [46]. 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 [47]. 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>82</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 [43], 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“ (Gründung in 2015 mit Sitz in Belgien) sein, diese europäische Strategie zu überdenken und abzugleichen mit anderen Initiativen und Standardisierungsbemühungen.

## Bewertung und Handlungsbedarf

Zunächst ist festzustellen, dass es eine Reihe sinnvoller Anwendungsfälle der Sekundärnutzung klinischer Daten für die klinische Forschung wie für die Versorgungsforschung gibt. Eine solche Sekundärnutzung ist dann auch ethisch und ökonomisch geboten, wenn sie unnötige Doppeluntersuchungen vermeiden hilft oder wissenschaftlich wichtige Studien erst ermöglicht bzw. deren Planung und Durchführung deutlich beschleunigt. Als unethisch sind beispielsweise klinische Studien anzusehen, die begonnen werden, obwohl eine Analyse vorhandener Daten hätte zeigen können, dass man die Rekrui-

---

82 siehe <https://www.i2b2.org/work/shrine.html>

tierungsziele 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. Leider ist völlig unklar, ob die aktuelle Initiative zur Schaffung einer europäischen Datenschutzverordnung [48] hier tatsächlich zu einer Verbesserung im Sinner einer Vereinheitlichung und Klärung beitragen kann. Die TMF hat hierzu in einer Stellungnahme konkrete Änderungsvorschläge vorgelegt.<sup>83</sup> Der in den derzeitigen Diskussionen zu dieser Gesetzesinitiative immer wieder auftauchende Vorschlag, den europäischen Rechtsrahmen für die Nutzung medizinischer Daten sehr eng zu ziehen und dafür auf nationaler Ebene Ausnahmeregelungen zuzulassen, wird gerade der uneinheitlichen deutschen Rechtsituation in keiner Weise gerecht.

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

---

83 siehe <https://www.tmf-ev.de/News/Stellungnahmen.aspx>

Versorgungs- als auch die Forschungsseite repräsentiert ist [21]. Die aktuelle Diskussion um ein eHealth-Gesetz bietet hierfür Chancen.

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 [19].

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 als vielversprechendste Lösung [23]. 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 [21].

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 betreiben. 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 solche forschungsunterstützenden Funktionen von den klinischen Einrichtungen vermehrt angefragt werden.

## 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 Inf Med*, 2014. **53**(6).
2. Kalra, D., Schmidt, A., Potts, H.W.W., Dupont, D., Sundgren, M., De Moor, G., *Case Report from the EHR4CR Project—A European Survey on Electronic Health Records Systems for Clinical Research*. *iHealth Connections*, 2011. **1**(2): S. 108–113
3. 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.
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. Prokosch, H.U., Beck, A., Ganslandt, T. et al., *IT Infrastructure Components for Biobanking*. Applied Clinical Informatics, 2010. **1**(4): S. 419–429.
14. 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.
15. Schneider, U.K., *Sekundärnutzung klinischer Daten - Rechtliche Rahmenbedingungen*. in Druck, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
16. BGH, *Bundesgerichtshof Beschluss VI ZR 135/13 vom 28. Oktober 2014 in dem Rechtsstreit Richtlinie 95/46/EG (Datenschutz-Richtlinie) Art. 2 a, Art. 7 f; TMG §§ 12, 15, 2014*, Bundesgerichtshof.
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 9.11.2013*. 2013. Arbeitskreis Medizinischer Ethik-Kommissionen, <http://www.ak-med-ethik-komm.de/formulare.html> (Abruf: 2013-12-11).
18. Pommerening, K., Drepper, J., Helbing, K., Ganslandt, T., *Leitfaden zum Datenschutz in medizinischen Forschungsprojekten - Generische Lösungen der TMF 2.0*. 2014, Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft.
19. 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).
20. 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.
21. 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.

22. 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.
23. 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: 2014-08-11).
24. Dugas, M., *Missing semantic annotation in databases. The root cause for data integration and migration problems in information systems*. *Methods Inf Med*, 2014. 53(6): S. 516–517.
25. Drepper, J., Semler, S.C., *Eine gemeinsame Sprache in der Medizin? Datenstandardisierung in medizinischer Versorgung und Forschung*, in *Telemedizinführer Deutschland*, Hrsg.: A. Jäckel. 2008: Bad Nauheim. S. 170–176.
26. Semler, S.C., *Labor- und Vitaldatenkommunikation: Standardisierung durch LOINC*. *Deutsches Ärzteblatt International*, 2003. 100(11): S. 10–13.
27. 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.
28. Geßner, C., *Kodierung von Einheiten physikalischer Messgrößen mit UCUM*. *HL7-Mitteilungen*, 2006(21): S. 6–17.
29. 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.
30. 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.
31. 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.
32. Ouagne, D., Hussain, S., Sadou, E., Jaulent, 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.
33. 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.

34. de Montjoie, A.J., *CDISC-Standards: reaching out to a global audience*. World Pharma Network, 2008: S. 10–12.
35. 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.
36. 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.
37. 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.
38. 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.
39. Schreiweis, B., Trinczek, B., Kopcke, 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.
40. 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.
41. 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.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.

47. 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.
48. EC Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr (Datenschutz-Grundverordnung). 2012. Europäische Kommission, [http://ec.europa.eu/justice/data-protection/document/review2012/com\\_2012\\_11\\_de.pdf](http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_de.pdf) (Abruf: 2014-06-10).



# Abkürzungsverzeichnis

AAL	Ambient Assisted Living
ACGT	Advancing Clinico Genomic Trials on Cancer ( <a href="http://www.eu-acgt.org">www.eu-acgt.org</a> )
ADT	Arbeitsgemeinschaft Deutscher Tumorzentren e.V. ( <a href="http://www.tumorzentren.de">www.tumorzentren.de</a> )
AG	Arbeitsgruppe
AGPLv3	GNU Affero General Public License Version 3
AID-Net	Autoinflammatory disorders in children and adolescence, im Rahmen des Förderschwerpunkts für seltene Erkrankungen vom BMBF gefördertes Forschungsnetzwerk
AIDS	Acquired Immune Deficiency Syndrome: Krankheitsbild einer erworbenen Immunschwäche aufgrund einer HIV-Infektion
AMG	Gesetz über den Verkehr mit Arzneimitteln – Arzneimittelgesetz
AMIA	American Medical Informatics Association ( <a href="http://www.amia.org">www.amia.org</a> )
ANALYZE	An der Mayo Clinic entwickeltes Format für im Rahmen der funktionellen Bildgebung gewonnene Bilddaten
APIS	Arztpraxisinformationssystem
ARCUS	Advanced Reader Certification for Unified Studies
ASCII	American Standard Code for Information Interchange; standardisierter Zeichensatz, der von den meisten Computersystemen interpretiert werden kann.
ASP	Application Service Provider
BAM	Compressed Binary SAM-Format
BCL	Base Call File
BDSG	Bundesdatenschutzgesetz
BGB	Bürgerliches Gesetzbuch
BGBI	Bundesgesetzblatt ( <a href="http://www.bgbl.de">www.bgbl.de</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> )
BioMedBridges	EU-Projekt zur Entwicklung gemeinsamer, harmonisierter Lösungen für die biomedizinischen ESFRI-Infrastrukturen
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> )
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> )
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> )

CD	Compact Disc
CDISC	Clinical Data Interchange Standards Consortium ( <a href="http://www.edisc.org">www.edisc.org</a> )
CDMS	Clinical Datamanagement System
CEN	Comité Européen de Normalisation, Europäisches Komitee für Normung ( <a href="http://www.cenorm.be">www.cenorm.be</a> )
CFR	Code of Federal Regulations der USA ( <a href="http://www.gpoaccess.gov/cfr">www.gpoaccess.gov/cfr</a> )
ChIP	Chromatin Immunoprecipitation (Sequencing)
COE	Council of Europe – Europarat ( <a href="http://www.coe.int">www.coe.int</a> )
COPD	Chronic Obstructive Pulmonary Disease - Chronisch obstruktive Lungenerkrankung
CRF	Case Report Form
CRM	Customer Relationship Management
CRO	Contract Research Organisation
CT	Computer-Tomografie
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> )
CURE-Net	Netzwerk für Congenitale Uro-Rektale Fehlbildungen ( <a href="http://www.cure-net.de">www.cure-net.de</a> )
DBR	Deutsches Biobanken-Register ( <a href="http://www.biobanken.de">www.biobanken.de</a> )
DFG	Deutsche Forschungsgemeinschaft ( <a href="http://www.dfg.de">www.dfg.de</a> )
DGU	Deutsche Gesellschaft für Unfallchirurgie e.V. ( <a href="http://www.dgu-online.de">www.dgu-online.de</a> )
DICOM	Digital Imaging and Communications in Medicine ( <a href="http://medical.nema.org">http://medical.nema.org</a> )
DKTK	Deutsches Konsortium für Translationale Krebsforschung ( <a href="http://www.dkfz.de/de/dktk">www.dkfz.de/de/dktk</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> )
DRKS	Deutsches Register Klinischer Studien ( <a href="http://www.drks.de">www.drks.de</a> )
DuD	Zeitschrift „Datenschutz und Datensicherheit“ ( <a href="http://www.dud.de">www.dud.de</a> )
DVD	Digital Versatile Disk; optisches Speichermedium
DVMD	Deutscher Verband Medizinischer Dokumentare ( <a href="http://www.dvmd.de">www.dvmd.de</a> )
e:Med	e:Med – Maßnahmen zur Etablierung der Systemmedizin; Forschungs- und Förderkonzept des BMBF
EC	European Commission
eCRF	electronic Case Report Form
ECRIN	European Clinical Research Infrastructures Network, seit 2007 als ESFRI-Projekt gefördert ( <a href="http://www.ecriin.org">www.ecriin.org</a> )
EDC	Electronic Data Capturing

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> )
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> )
EN	Europäische Norm des CEN
EORTC	European Organisation for Research and Treatment of Cancer ( <a href="http://www.eortc.be">www.eortc.be</a> )
EP	Europäisches Parlament ( <a href="http://www.europarl.europa.eu">www.europarl.europa.eu</a> )
E-PIX	Im Rahmen des GANI_MED-Projekts entwickelte und mit dem PIX-Profilen von IHE kompatible MPI-Software
ePRO	electronic Patient Reported Outcomes
ERIC	European Research Infrastructure Consortium; europäisches Rechtsinstrument für Forschungsinfrastrukturen
ESBB	European, Middle Eastern & African Society for Biopreservation and Biobanking; Unterorganisation der ISBER ( <a href="http://www.esbb.org">www.esbb.org</a> )
eSDI	electronic Source Data Interchange
ESFRI	European Strategy Forum on Research Infrastructures ( <a href="http://cordis.europa.eu/esfri">http://cordis.europa.eu/esfri</a> )
eTMF	electronic Trial Master File
eTRIKS	European Translational Information and Knowledge Management Services ( <a href="http://www.etriks.org">www.etriks.org</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> )
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> )
FASTQ	Textbasiertes Format zur Speicherung von Sequenz- und Qualitätsdaten
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
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> )
GB	Gigabyte(s)
GCP	Good Clinical Practice, Regelwerk der ICH
GEKID	Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. ( <a href="http://www.gekid.de">www.gekid.de</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> )
GG	Grundgesetz der Bundesrepublik Deutschland
GI	Gesellschaft für Informatik e.V. ( <a href="http://www.gi-ev.de">www.gi-ev.de</a> )
GMDS	Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. ( <a href="http://www.gmds.de">www.gmds.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> )
GOLDnet	Deutsches Netzwerk für diffus parenchymatöse Lungenerkrankungen ( <a href="http://www.dpld.de">www.dpld.de</a> )
gPAS	generic Pseudonym Administration Service
GPS	Global Positioning System
GSP	Good Scientific Practice
GTDS	Gießener Tumordokumentations-System ( <a href="http://www.gtds.de">www.gtds.de</a> )
GVF	Genome Variation Format
HCV	Hepatitis-C-Virus
HEGP	Hôpital Européen Georges-Pompidou ( <a href="http://www.hegp.fr">www.hegp.fr</a> )
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> )
HPLC-MS	High-Performance Liquid Chromatography – Mass Spectrometry
HTTP	Hyper Text Transfer Protocol
i2b2	Informatics for Integrating Biology and the Bedside ( <a href="http://www.i2b2.org">www.i2b2.org</a> )
IBE	Institut für medizinische Informationsverarbeitung, Biometrie und Epidemiologie der Uni München ( <a href="http://ibe.web.med.uni-muenchen.de">http://ibe.web.med.uni-muenchen.de</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> )
ICT	Information Communication Technology
ID	Identifikationsnummer
IFB	Integrierte Forschungs- und Behandlungszentren; Fördermaßnahme des BMBF
IHE	Integrating the Healthcare Enterprise ( <a href="http://www.ihe.net">www.ihe.net</a> )
IIT	Investigator initiated trial
IMI	Innovative Medicines Initiative ( <a href="http://www.imi-europe.org">www.imi-europe.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> )
INFOPAT	Informationstechnologie für patientenorientierte Gesundheitsversorgung in der Metropolregion Rhein-Neckar
ISBER	International Society for Biological and Environmental Repositories ( <a href="http://www.isber.org">www.isber.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> )
IZKS	Interdisziplinäres Zentrum Klinische Studien
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> )
JPEG2000	Komprimiertes Bilddateiformat (ISO 15444) der Joint Photographic Experts Group ( <a href="http://www.jpeg.org">www.jpeg.org</a> ); Nachfolger des JPEG-Formats
JPIP	JPEG2000 Interactive Protocol
k-Anonymisierung	Verfahren zur Anonymisierung einer Datensammlung, so dass jede Merkmalskombination, die potentiell für einen reidentifizierenden Abgleich genutzt werden könnte, in mindestens k Datensätzen vorkommt
KIS	Krankenhausinformationssystem
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> )
KKS	Koordinierungszentrum für Klinische Studien ( <a href="http://www.kks-netzwerk.de">www.kks-netzwerk.de</a> )
KKSN	Netzwerk der Koordinierungszentren 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> )

KoRegIT	TMF-Projekt zur Erstellung eines Anforderungskatalogs zur IT-Unterstützung von Kohorten und Registern
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> )
LAMP	Open-Source-Plattform für Webanwendungen unter Verwendung von Linux, Apache, MySQL und PHP
LC-MS	Liquid Chromatography – Mass Spectrometry
LIMS	Laboratory Information Management System
Mainzliste	Webbasierte Open-Source-Pseudonymisierungsplattform der Universitätsmedizin Mainz ( <a href="http://www.unimedizin-mainz.de/imbei/informatik/opensource/mainzliste.html">http://www.unimedizin-mainz.de/imbei/informatik/opensource/mainzliste.html</a> )
MB	Megabyte(s)
MD	Muskeldystrophie
mdi	mdi – Forum der Medizin_Dokumentation und Medizin_Informatik; von den beiden Verbänden BVMI und DVMD gemeinsam herausgegebene Fachzeitschrift
MD-NET	Muskeldystrophie-Netz ( <a href="http://www.md-net.org">www.md-net.org</a> )
MDPE	Medical Data and Picture Exchange: Im KN POH entwickelte Teleradiologielösung auf Basis des Modells A der generischen Datenschutzkonzepte der TMF.
MDR	Metadata Repository
MethInfraNet	Maßnahmen zur methodischen und infrastrukturellen Vernetzung für Qualitäts- und Effizienzsteigerung in der medizinischen Forschung; BMBF-Zuwendung für Ausbau und Verstetigung der TMF
MEVIS	Fraunhofer-Institut für Bildgestützte Medizin ( <a href="http://www.mevis.fraunhofer.de">www.mevis.fraunhofer.de</a> )
MI	Medizinische Informatik
MIAME	Minimum Information About a Microarray Experiment
MIAPE	Minimum Information About a Proteomics Experiment
MIGS	Minimum Information about a (Meta)Genome Sequence
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> )
MO	Missouri, Staat in den USA
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> )

MPG	Gesetz über Medizinprodukte - Medizinproduktegesetz
MPI	Master Patient Index
MRI	Magnetic Resonance Imaging
mRNA	messenger-RNA
MRT	Magnetresonanztomographie
MS	Massen-Spektrometrie
MSI	Metabolomics Standards Initiative
MTA	Medizinisch-Technische(r) Assistent(in)
NatKo	Nationale Kohorte ( <a href="http://www.nationale-kohorte.de">www.nationale-kohorte.de</a> )
NCI	US National Cancer Institute ( <a href="http://www.cancer.gov">www.cancer.gov</a> )
NER	Nationale Ethikrat ( <a href="http://www.ethikrat.org">www.ethikrat.org</a> )
NGFN	Nationales Genomforschungsnetz, vom BMBF gefördert ( <a href="http://www.ngfn.de">www.ngfn.de</a> )
NGS	Next-Generation Sequencing
NIfTI	Neuroimaging Informatics Technology Initiative ( <a href="http://nifti.nih.gov">http://nifti.nih.gov</a> )
NIRK	Network for Ichthyoses and Related Keratinization Disorders ( <a href="http://www.netzwerk-ichthyose.de">www.netzwerk-ichthyose.de</a> )
OASIS	Organization for the Advancement of Structured Information Standards ( <a href="http://www.oasis-open.org">www.oasis-open.org</a> )
OME	Open Microscopy Environment Consortium ( <a href="http://www.openmicroscopy.org">www.openmicroscopy.org</a> )
OpenClinica	Open-Source-Software für klinische Studien mit Studiendatenmanagement- und EDC-Funktionen ( <a href="http://www.openclinica.org">www.openclinica.org</a> )
OSSE	Open-Source-Registersystem für Seltene Erkrankungen in der EU; vom BMG gefördertes Projekt im Rahmen des Aktionsplans für Menschen mit seltenen Erkrankungen
PACS	Picture Archiving and Communication System
PB	Petabyte
PC	Personal Computer
PDA	Personal Digital Assistant
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
PET	Positronenemissionstomographie
PHP	Rekursives Akronym für PHP Hypertext Preprocessor: Programmiersprache zur Erstellung von Webseiten
PID	Patientenidentifikator
PIX	Patient Identifier Cross-referencing, IHE-Profil zum domänenübergreifenden Abgleich von Patienten-Identifikatoren

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> )
PRO	Patient-Reported Outcome
QM	Qualitätsmanagement
QuaSi-Niere	Qualitätssicherung in der Nierenersatztherapie – QuaSi-Niere gGmbH ( <a href="http://www.quasi-niere.de">www.quasi-niere.de</a> )
RDA	Research Data Alliance ( <a href="https://rd-alliance.org">https://rd-alliance.org</a> )
RIS	Radiologie-Informations-System
RNA	Ribonukleinsäure
RNA-SEQ	Gesamt-Transkriptom-Shotgun-Sequenzierung
RöV	Verordnung über den Schutz vor Schäden durch Röntgenstrahlen – Röntgenverordnung
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
SaaS	Software as a Service
SAE	Serious Adverse Event, schwerwiegendes unerwünschtes Ereignis im Rahmen einer Arzneimittelprüfung
SAM	Sequence Alignment/Map, generisches Format zur Speicherung von Nukleotid-Sequenzen
SCTO	Swiss Clinical Trial Organisation ( <a href="http://www.scto.ch">www.scto.ch</a> )
SDO	Standards Development Organization
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> )
SGB	Sozialgesetzbuch
SHARE	CDISC Shared Health and Research Electronic Library
SHIP	Study of Health in Pomerania: Vom BMBF geförderte Bevölkerungsgesundheitsstudie ( <a href="http://ship.community-medicine.de">http://ship.community-medicine.de</a> )
SNP	Single Nucleotide Polymorphism, Variationen von einzelnen Basenpaaren in einem DNA-Strang
SNV	Single Nucleotide Variant
SOAP	Simple Object Access Protocol; vom W3C empfohlener, XML-basierter Protokoll-Standard zur Kommunikation strukturierter Daten mit Webservices per HTTP

SOP	Standard Operating Procedure
SPM	Statistical Parametric Mapping; Open-Source-Software auf der Basis von MATLAB zur Analyse von Bilddaten aus der funktionellen Bildgebung
SRA	Sequence Read Archive Schema des European Nucleotide Archive
sysINFLAME	Network for Systemic Medicine for Chronic Inflammatory Diseases; im Rahmen der e:MED-Förderlinie des BMBF gefördertes Projekt
TAILOR	The Alliance for Interoperability in IT for JOint Medical Research Infrastructures; Antrag auf eine Förderung im Rahmen der Methodenförderung zur individualisierten Medizin des BMBF
TB	Terabyte
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> )
TranSarNet	Translational Sarcoma Research Network, im Rahmen des Förderschwerpunkts für seltene Erkrankungen vom BMBF gefördertes Forschungsnetzwerk ( <a href="http://campus.uni-muenster.de/transarnet.html">http://campus.uni-muenster.de/transarnet.html</a> )
tranSMART	Webbasierte Knowledge-Management-Plattform für die wissenschaftliche Hypothesenentwicklung auf Basis von Beziehungen zwischen phänotypischen und genetischen Daten ( <a href="http://go.transmart.etriks.org">http://go.transmart.etriks.org</a> )
TTP	Trusted Third Party
UC	Use Case
UMTS	Universal Mobile Telecommunications System, Mobilfunkstandard
VCF	Variant Call Format, Textdatei-Format für die Speicherung von Informationen zu Genompositionen und zugehörigen Metadaten
VISTA	Webbasierte Studiensoftware der EORTC
VPH	Virtual Physiological Human
W3C	World Wide Web Consortium ( <a href="http://www.w3.org">www.w3.org</a> )
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
XDS	Cross Enterprise Document Sharing, IHE-Integrationsprofil
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> )
ZKS	Zentrum für Klinische Studien



Der Report des IT-Reviewing-Boards zu „IT-Infrastrukturen in der patientenorientierten Forschung in Deutschland“ wird jährlich mit wechselnden Schwerpunkten, Themen und Autoren fortgeschrieben. Schwerpunkt der ersten Ausgabe war die Datenerhebung in der klinischen Forschung. Die aktuelle zweite Ausgabe schreibt diesen Schwerpunkt fort, enthält jedoch erstmals auch jeweils ein Kapitel zur Erhebung und Verarbeitung molekularbiologischer Daten sowie zur Sekundärnutzung klinischer Daten für die Forschung. Alle weiteren Kapitel wurden umfassend aktualisiert. Das Board setzt sich aus Experten für IT-Infrastrukturen in der medizinischen Forschung und Versorgung zusammen und ist bei der TMF als Dachorganisation der medizinischen Verbundforschung in Deutschland angesiedelt.

Die TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. ist die Dachorganisation der medizinischen Verbundforschung in Deutschland. Sie 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. Damit wird die Sicherheit, Qualität und Effizienz der medizinischen Forschung gesteigert, für die notwendige Standardisierung gesorgt und Forschern, unabhängig von ihrer wissenschaftlichen Fragestellung, Hilfestellung bei der Bewältigung der zunehmend komplexer werdenden gesetzlichen und regulatorischen Rahmenbedingungen angeboten.

ISSN 2363-5754

ISBN 978-3-89838-701-9



9 783898 387019

AKA

[www.aka-verlag.com](http://www.aka-verlag.com)

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

