

IT-Infrastrukturen in der patientenorientierten Forschung

Aktueller Stand und Handlungsbedarf

2015

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 – 2015

Redaktion & Herausgabe

Dr. Johannes Drepper, Sebastian C. Semler

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

Bibliografische Information der Deutschen Nationalbibliothek

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

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

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

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

Satz: Anja Brysch, Michael Pohl

Verlag: Akademische Verlagsgesellschaft AKA GmbH,
Postfach 22 01 16, 14061 Berlin | www.aka-verlag.com

Druck: CPI buch bücher.de gmbH, Birkach

Printed in Germany

ISSN: 2363-5754

ISBN: 978-3-89838-710-1

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.

In den ersten beiden Berichten sind die verschiedenen Datenquellen in den Blick genommen worden: von klinischen Studien über Register und Kohorten bis zur Omics-Technologie und der sekundären Nutzung klinischer Behandlungsdaten. Für jeden Bereich wurde dargestellt, welche Lösungen möglich sind. Als Querschnittsaufgabe ist bereits das ID-Management beschrieben worden, also die Generierung und Verwaltung eindeutiger Kennzeichner für alle Probanden, Datensätze und Proben. Diese Darstellung hat allerdings in Form einer Leerstelle auf das Problem der Integration dieser heterogenen Daten verwiesen. Wie können die verschiedenen Datenarten zueinander in Beziehung gesetzt und übergreifend ausgewertet werden? Wie eine IT-Architektur hierfür gestaltet sein muss und welche Softwarelösungen notwendig sind, das wird nun in diesem und den folgenden Berichten einen Schwerpunkt der Untersuchung bilden.

Während es für einzelne Datenarten und Forschungsprojekte an ausgewählten Standorten schon sehr gute IT-Lösungen gibt, scheint die Integration immer noch eine große Herausforderung darzustellen. Dies gilt sowohl standortbezogen als auch in Forschungsverbänden und damit standortübergreifend. Vor diesem Hintergrund werden für die weitere Entwicklung der IT-Infrastrukturen in der medizinischen Forschung in Deutschland die folgenden Anforderungen und Empfehlungen gesehen:

- ▶ **Sichtbarkeit und Transparenz:** Die Forderung nach einer systematischen Übersicht über bestehende Systeme, Ressourcen und Datenbestände in der medizinischen Forschung zieht sich durch alle im IT-Report angesprochenen Themenfelder. Initiativen hierzu sind in den vergangenen Jahren sowohl aus der TMF-Community hervorgegangen als auch von außerhalb gestartet worden. Allerdings besteht nach wie vor die Schwierigkeit, solche Infrastrukturen aus einer Projektförderung in den langfristigen Betrieb zu überführen. Hier sind geeignete Finanzierungs- und Betriebsmodelle sowie Anreizsysteme zu entwickeln, insbesondere um diese Angebote in einem dynamischen Wissenschaftsumfeld laufend aktuell zu halten. Dies kann nur mit Unterstützung der Politik und der Förderorganisationen geschehen.

- ▶ **Standardisierung und Interoperabilität:** Medizinischer Fortschritt beruht immer stärker auf der Möglichkeit, Informationen aus verschiedenen Sektoren und IT-Systemen zusammenzuführen. Die Nutzung von Standards ist daher für die Erfassung, Verarbeitung, Integration und Weitergabe von Daten unerlässlich. Neben der Harmonisierung der Daten- und Übertragungsstandards ist insbesondere auch eine standardisierte 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.
- ▶ **Internationale Anschlussfähigkeit:** Medizinische Forschung geschieht heute sowohl eingebettet in einen weltweiten Informationsaustausch als auch im internationalen Wettbewerb stehend. Gerade die dynamische Methodenentwicklung im Bereich der Omics-Technologien zeigt, dass deutsche Forschungsstandorte von internationalen Tools, Methoden und Standards nicht nur profitieren können, sondern auch müssen. Unnötige Eigenentwicklungen und unabgestimmte Alleingänge im methodischen Bereich sind unökonomisch und führen schlimmstenfalls zu qualitativ nicht ausreichend abgesicherter Forschung sowie fehlender internationaler Anschlussfähigkeit. Die von der TMF unterstützte Evaluation, Weiterentwicklung und Dissemination der aus Boston stammenden Datenplattform für translationale Forschung i2b2 kann hier als positives Beispiel gezählt werden.
- ▶ **Aufbau von Dienstleistungsangeboten:** Vermehrt wird der Bedarf geäußert, im Bereich der akademischen medizinischen Forschung zentral angebotene Dienstleistungen zu moderaten Preisen zu etablieren. Dies gilt von der Nutzung einzelner Softwarekomponenten über das Angebot eines Identitätsmanagements im Rahmen der benötigten Datenschutzkonzepte bis hin zu einer deutschen Cloud für „Applied & Translational Genomics“. Bedarfsermittlung, Planung und Aufbau solcher notwendigerweise standardisierten Services können nur in Standort-übergreifender

Abstimmung geschehen. Notwendig für die Koordination ist eine neutrale und außerhalb des wissenschaftlichen Wettbewerbs stehende Stelle.

- ▶ **Klärung rechtlicher Fragen:** Mit der Weiterentwicklung des medizinischen Wissens und der technologischen Möglichkeiten sind auch weiterhin neue rechtliche und ethische Fragen zu beantworten. Das gilt unter anderem für den zunehmenden Einsatz mobiler Technologien in medizinischer Forschung und Versorgung, im Bereich der Sekundärnutzung von Versorgungsdaten für Forschung und Qualitätssicherung sowie nicht zuletzt auch für den Umgang mit genetischen Daten aus Hochdurchsatz-Analysen (Omics-Technologien).
- ▶ **Finanzierung:** IT-Infrastrukturen in der medizinischen Forschung sind nach wie vor nicht ausreichend finanziert. IT-Systeme werden häufig im Rahmen von zeitlich begrenzten (Verbund-)Projekten aufgebaut, weshalb sie oft zu wenig in die Gesamtstruktur einer Einrichtung eingebettet sind und nach Ablauf des Förderzeitraumes nicht mehr optimal betrieben werden können. Besonders problematisch ist auch, dass qualifiziertes Personal, das über das Wissen und die Erfahrung im Umgang mit den IT-Systemen verfügt, angesichts von Zeitverträgen im Rahmen von Projekten oftmals nicht gehalten werden kann. Das aktuell vom BMBF ausgedescribte Förderkonzept „Medizininformatik“ wird vor diesem Hintergrund als positiver Impuls gewertet.

Neue Themenschwerpunkte 2015/2016

Forschungsdatenmanagement

Während in den vorherigen Ausgaben des IT-Reports vorwiegend die einzelnen Datenquellen beschrieben wurden, widmet sich die vorliegende Ausgabe explizit erstmals auch dem Querschnittsthema des integrativen Forschungsdatenmanagements. Die Vokabel Forschungsdatenmanagement hat aufgrund ihrer Bedeutung mittlerweile Eingang in die Förderkonzepte nationaler und internationaler Drittmittelgeber gefunden. Hintergrund ist die Erkenntnis, dass viele bislang als qualitativ hochwertig angesehene wissenschaftliche Arbeiten bei genauerer Untersuchung nicht nachvollzogen werden können. Ein belastbares – jedoch durchaus mit Aufwand verbundenes – Forschungsdatenmanagement hat daher zur Aufgabe, jegliche Datenquelle mit klarer Herkunft, Zeitstempel, Autorenschaft etc. dauerhaft zu belegen.

Die ganze Verarbeitungskette von der Extraktion der verschiedenen Daten aus den jeweiligen Quellsystemen bis zur Analyse und Auswertung der Daten kann im biomedizinischen Umfeld mit der Kombination verschiedener Werkzeuge erreicht werden. Allerdings können viele der in der Verbundforschung hierfür eingesetzten Werkzeug-

ge noch nicht die erforderliche Datenqualität und Reproduzierbarkeit von Analyseergebnissen garantieren. Für kritische Analysen schränkt dies derzeit die Auswahl geeigneter Werkzeuge noch deutlich ein.

Trotzdem sollten Eigenentwicklungen soweit möglich vermieden und eher der Anschluss an internationale Entwicklungen gesucht werden. Dies erfordert im Sinne der Ökonomie und der guten wissenschaftlichen Praxis dringend eine bessere Abstimmung bzw. einen intensiveren Austausch zwischen Forschungsprojekten. Hierfür wird die Nutzung bewährter Plattformen wie der AG IT-Infrastruktur und Qualitätsmanagement der TMF empfohlen. Auch in Bezug auf die Einführung und Abstimmung von Regeln, die die Rahmenbedingungen für die Speicherung, Weitergabe und Analyse der verschiedenen Daten im Forschungsdatenmanagement transparent festlegen, wird ein intensiverer Austausch für notwendig gehalten.

Metadaten und Metadata Repositories

Die Verwertbarkeit von Primär- und Sekundärdaten der biomedizinischen Forschung hängt entscheidend von der Qualität und der Verfügbarkeit der sie beschreibenden Metadaten ab. Die Notwendigkeit einheitlicher und detaillierter Definitionen von Datenelementen und Formularen ist dabei im Hinblick auf eine qualitativ hochwertige Dokumentation anerkannt. Der Aufwand für die Erstellung eines guten Dokumentationskonzepts ist jedoch hoch und erfordert die interdisziplinäre Zusammenarbeit von klinischen Forschern, Statistikern und IT-Experten. Gleiches gilt für Formulare und elektronische Masken in Informationssystemen der Versorgung. Daher wird die Verwendung standardisierter Metadatenmodelle, elemente und kataloge empfohlen.

Während die theoretische Fundierung von Metadatenmodellen und -systemen weit fortgeschritten ist, fehlt es noch an reifen Werkzeugen, die die Forscher mit quantifizierbarem Mehrwert bei ihren Arbeitsprozessen unterstützen. Zu oft werden noch projektspezifische Insellösungen entwickelt. Daraus resultieren starke Verzögerungen bei der Bereitstellung von Forschungsprimärdaten für Auswertungen und bei der Zusammenführung verschiedener Datensätze. Auch die Standardisierung von Inhalten kommt nur langsam voran – Definition und Harmonisierung von Datenelementen sind selten ein prioritäres Anliegen der biomedizinischen Forscher.

Anonymisierung von medizinischen Individualdaten

Das Ziel von Anonymisierungsmethoden besteht darin, durch eine Veränderung personenbezogener Merkmale in einem Datensatz die Identität der Patienten oder Probanden zu

verbergen und somit Missbrauch zu verhindern. Mit Methoden der Anonymisierung ist es möglich, medizinische Forschungsdaten außerhalb der Bestimmungen des Datenschutzrechts zu verarbeiten. Im rechtlichen Sinne handelt es sich bei anonymisierten Daten nicht mehr um personenbezogene Daten. Dies unterscheidet die Anonymisierung wesentlich von der Pseudonymisierung. Daher ist Anonymisierung vor allem für „Data Sharing“ und Sekundärnutzung relevant. Wesentliche Beispiele sind der intra- und inter-institutionelle Austausch von Daten für Forschungszwecke und die Nutzung von im Behandlungskontext gesammelten klinischen Daten für die Forschung.

Anonymisierungsmethoden haben das Potential, neue Anwendungsfälle zu erschließen, etablierte Schutzmaßnahmen zu komplementieren und datenintensive Prozesse sicherer zu machen. Allerdings fehlt es noch an praktischen Erfahrungen, welche Anwendungsfälle auf Basis anonymisierter und damit häufig vergrößerter Daten durchgeführt werden können und welche eher nicht. Nötig sind darüber hinaus abgestimmte und in einem Leitfaden dokumentierte Verfahren zur Anonymisierung, die zu einer breiteren Anwendung dieser Methoden führen und gleichzeitig den Anwendern auch die notwendige Sicherheit geben, datenschutzkonform zu handeln. Dabei müssen ggf. auch die mathematischen und statistischen Methoden der Anonymisierung mit anderen Sicherheits- und Schutzmaßnahmen, beispielsweise organisatorischer Art, kombiniert werden.

Cloud-Computing in der medizinischen Forschung

Die Nutzung von Cloud-Computing wird heute als wesentlicher IT-Trend angesehen, der eine hohe Flexibilität und Kosteneffizienz ermöglicht und daher auch Auswirkungen auf IT-Infrastrukturen in der Forschung hat. Dies gilt in besonderem Maße für die biomedizinische Forschung, die in den letzten Jahren von einem starken Anstieg der Anforderungen an Speicher- und Rechenkapazität gekennzeichnet ist. Technisch gesehen ist die Cloud ein virtualisiertes Angebot von IT-Services. Dabei können Hardwareressourcen, im Wesentlichen Rechenkapazität und Speicherplatz, Anwendungs- und Entwicklungsplattformen als auch fertige Anwendungen über Netzwerkverbindungen angeboten werden.

Grundsätzlich erscheint eine Weiterentwicklung der IT-Infrastrukturen in der medizinischen Forschung hin zu einer stärkeren Service-Orientierung wünschenswert. Hierfür spricht die in Deutschland auf sehr viele Standorte verteilte Forschungsaktivität. Dabei werden an vielen einzelnen Standorten, nicht selten auch noch von mehreren Abteilungen unabhängig voneinander, Infrastrukturen für die Forschung aufgebaut. Es ist aber nicht davon auszugehen, dass alle diese Einrichtungen und Abteilungen, zumal bei stark begrenzten Budgets, eine allen Qualitätsanforderungen genügende

Forschungs-IT entwickeln und betreiben können. Eine Reihe deutscher Forscher kommt in einem White Paper¹ zur Entwicklung einer deutschen Cloud für „Applied & Translational Genomics“ zu dem Schluss, dass keine einzige deutsche Universität oder Forschungseinrichtung derzeit über die notwendige Infrastruktur verfügt, um mit den absehbaren Datenmengen im Omics-Bereich Analysen durchzuführen und eine gesicherte Speicherung sowie langfristigen Zugriff auf die Daten zu gewährleisten. Daher werden zentralisierte und auf Cloud-Technologien basierende Angebote perspektivisch auch in der medizinischen Forschung vermehrt Anwendung finden. Grundsätzlich können solche Angebote rechtssicher aufgebaut und betrieben werden. Bestimmte Einschränkungen und Auflagen sind dabei allerdings zu beachten.

Strategische Aktivitäten und Fördermaßnahmen im Bereich der IT-Infrastrukturen

Ein neues Kapitel zu strategischen Aktivitäten und Fördermaßnahmen im Bereich der IT-Infrastrukturen schließt den vorliegenden Report ab. Hier sind erstmals die zahlreichen Initiativen, Gremien, Stellungnahmen, Empfehlungen, Leitlinien und Roadmaps zu Forschungsinfrastrukturen mit IT-Bezug nebeneinander dargestellt. So werden Überlappungen, unterschiedliche Akzent- und Schwerpunktsetzungen sowie alternative Lösungsansätze erkennbar. Zusammen mit einer Übersicht der Fördermaßnahmen in diesem Bereich macht dieses Kapitel die Vielfältigkeit deutlich, mit der auf nationaler wie europäischer Ebene das zentrale Thema der IT-Infrastrukturen in der medizinischen Forschung behandelt wird.

Entwicklungen und Fortschritte seit 2014

- ▶ Die TMF-finanzierte Projektserie zur Entwicklung eines Integrated Data Repository Toolkits auf Basis der Open-Source-Plattform i2b2 konnte 2015 abgeschlossen werden. Damit steht eine international anschlussfähige und an deutsche Rahmenbedingungen (Schnittstellen, Datenschutz) angepasste Tool-Sammlung für bestimmte Aufgaben im Forschungsdatenmanagement samt Community-Unterstützung zur Verfügung.
- ▶ Für das im vorliegenden Report enthaltene Kapitel zu Registern und Kohorten ist eine erste Übersicht über Softwareangebote für das Datenmanagement in Registern erstellt worden. Bislang gab es viele Anfragen in den Gremien der TMF nach entsprechenden Lösungen, aber keine Übersicht über diesen sehr spezialisierten Markt mit eher kleineren Softwareanbietern.

¹ Brors, B., Eberhardt, W., Eils, R. et al. White paper: The Applied & Translational Genomics Cloud (ATGC). 2015, <http://www.applied-translational-genomics-cloud.de/joomla/index.php/de/download> (Abruf: 2016-01-25).

- ▶ Die Firma Apple hat im März 2015 mit dem ResearchKit ein Framework für die Erstellung Smartphone- oder Tablet-basierter Umfragen oder Studien vorgestellt. Das Tempo, mit dem dieses Framework in der internationalen Community bisher angenommen wurde, spricht dafür, dass diese Entwicklung die Landschaft der mobilen Datenerfassung in der medizinischen Forschung nachhaltig verändern wird.
- ▶ Einen nachhaltigen Einfluss auf die Entwicklung der Medizininformatik in Deutschland und ihren Einfluss auf die Verzahnung der Versorgungs- und Forschungs-IT wird das Förderkonzept „Medizininformatik“ des BMBF haben, welches im November 2015 veröffentlicht wurde. Ausgewählte Konsortien sollen in einer ersten Phase beispielhafte und später breit anwendbare Konzepte für die Integration verschiedener Datenbestände entwickeln. Von diesen Konzepten werden dann für die zweite Phase einige zur Umsetzung ausgewählt. Eine zentrale und unabhängige Begleitstruktur ist vorgesehen.
- ▶ Die DFG hat im Sommer 2015 eine Ausschreibung zur Förderung von Forschungsprojekten über die und mit der TMF veröffentlicht und im September 2015 zu einem Workshop zur Präsentation und Diskussion von Projektideen eingeladen. Daraus sind ca. 20 Projektanträge zur Weiterentwicklung der methodischen Unterstützung medizinischer Forscher hervorgegangen und bei der DFG eingereicht worden. Die Begutachtung findet Anfang 2016 statt. Es ist davon auszugehen, dass auch diese Maßnahme der DFG einen wichtigen Beitrag zur Weiterentwicklung der IT-Infrastrukturen in der medizinischen Forschung leisten wird.
- ▶ Die Verhandlung der EU-Datenschutzgrundverordnung zwischen den zuständigen EU-Gremien ist im Dezember 2015 abgeschlossen worden. Damit ist erstmals weitgehend absehbar, welche Neuerungen in Bezug auf datenschutzrechtliche Anforderungen auf die medizinische Forschung langfristig zukommen. Allerdings lässt der jetzt bekannte Verordnungstext weiterhin Spielräume für nationale Regelungen, gerade auch im Umgang mit Gesundheitsdaten, so dass eine abschließende Beurteilung noch nicht möglich ist.
- ▶ Die TMF hat 2015 ein umfangreiches Rechtsgutachten zu den rechtlichen Rahmenbedingungen der Sekundärnutzung klinischer Daten publiziert. In dem Buch der TMF-Schriftenreihe wird dabei erstmals auch im Detail auf alle hierfür relevanten Regelungen aller 16 Bundesländer und die Datenschutzregelungen der Kirchen eingegangen. Komplettiert wird dies Angebot durch ein Online-Tool, welches für jede Art von Krankenhaus in jedem Bundesland und für jeden Verwendungszweck schnell die relevanten Rechtsgrundlagen zusammenstellt.