

Aus dem Institut für Medizinische Biometrie und Informatik der Universität Heidelberg
Geschäftsführender Direktor: Prof. Dr. sc. hum. Meinhard Kieser

LEITFADEN ZUR INTEGRATION EINES KLINISCHEN
REGISTERS IN EIN EXISTIERENDES INFORMATIONSSYSTEM
AM BEISPIEL DES KLINISCHEN MYELOMREGISTERS
DES UNIVERSITÄTSKLINIKUMS HEIDELBERG

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht-Karls-Universität

vorgelegt von
MARTIN LÖPPRICH
aus
Hermannstadt (Rumänien)
2016

Dekan: Prof. Dr. med. Wolfgang Herzog
Doktormutter: Frau Prof. Dr. sc. hum. Petra Knaup-Gregori

INHALTSVERZEICHNIS

ABBILDUNGSVERZEICHNIS	v
TABELLENVERZEICHNIS	v
ABKÜRZUNGSVERZEICHNIS	vi
1 EINLEITUNG	1
2 GRUNDLAGEN	5
2.1 Klinische Register	5
2.2 Systematische Planung	7
2.3 Integration	8
3 ANALYSE EXISTIERENDER PLANUNGSINSTRUMENTE	11
3.1 Literaturrecherche	11
3.2 Memorandum Register für die Versorgungsforschung . .	12
3.3 Registries for Evaluating Patient Outcomes: A User's Guide	15
3.4 Systematic Planning of Clinical Documentation	18
3.5 Vergleich und Bewertung	21
3.6 Abgeleitete Anforderungen an einen Leitfaden	26
4 LEITFADEN FÜR EIN KLINISCHES REGISTER	29
4.1 Vorgehensmodell zur systematischen Planung	29
4.1.1 Beschreibung der Inhalte	30
4.1.2 Hinweise zur Anwendung	44
4.2 Referenz-IT-Architektur	45
4.2.1 Anwendungssysteme	46
4.2.2 Prolektives Dokumentationssystem	48
4.2.3 Integration mit Systemen der Versorgung	68
4.2.4 Integration mit Systemen der Forschung	78
4.2.5 Data Warehouse	86
4.2.6 Metadatenmanagementsystem	91
4.2.7 Hinweise zur Anwendung	96
5 ANWENDUNG DES LEITFADENS AM BEISPIEL DES KLINISCHEN MYELOMREGISTERS	99
5.1 Registerprotokoll klinisches Myelomregister	99
5.2 Bewertung elektronischer Datenerfassungssoftware . . .	117
5.3 Integration strukturierter Informationen	118
5.4 Integration unstrukturierter Informationen	121
5.5 Metadatenmanagementsystem klinisches Myelomregister	125
5.6 Abschließende Bemerkungen	127

6	DISKUSSION	129
6.1	Vorgehensmodell zur systematischen Planung	130
6.2	Referenz-IT-Architektur	132
6.3	Ausblick	134
7	ZUSAMMENFASSUNG	137
	LITERATUR	139
	EIGENE VERÖFFENTLICHUNGEN	147
	Anhang	149
A	TESTPROTOKOLL DYNAMISCHES DOKUMENTATIONSMODELL	151
B	SYSTEMBEWERTUNG OPENCLINICA UND ONKOSTAR	157
	LEBENS LAUF	161
	DANKSAGUNG	163

ABBILDUNGSVERZEICHNIS

Abbildung 1	Referenz-IT-Architektur	46
Abbildung 2	Gegenüberstellung Dokumentationsmodelle	49
Abbildung 3	Gegenüberstellung Client-Server-Modelle	63
Abbildung 4	Präsentationsintegration	68
Abbildung 5	Transkription über Text Mining	75
Abbildung 6	Transkription über Schnittstellen	76
Abbildung 7	Transkription über Exportfunktionen	77
Abbildung 8	Vertikale Integration	81
Abbildung 9	Horizontale Integration	82
Abbildung 10	Konzeptionelle Beziehung von Metadaten	93
Abbildung 11	Beziehung zwischen Merkmal und Wertebereich	94
Abbildung 12	Physisches Metadatenmodell	94
Abbildung 13	Beispiel physisches Metadatenmodell	95
Abbildung 14	Entscheidungs- und Realisierungsmöglichkeiten der Referenz-IT-Architektur	98
Abbildung 15	Präsentationslayout Progression	109
Abbildung 16	Präsentationslayout Gesamtüberleben	109
Abbildung 17	Erhebungsformulare Patient und Chemotherapie	114
Abbildung 18	Framework Text Mining	123
Abbildung 19	Relationales Datenmodell MDMS	126

TABELLENVERZEICHNIS

Tabelle 1	<i>PubMed</i> -Suchanfrage	11
Tabelle 2	Gliederung <i>Memorandum</i>	13
Tabelle 3	Gliederung <i>User's Guide</i>	16
Tabelle 4	Gliederung <i>Dokumentationsprotokoll</i>	20
Tabelle 5	Vergleich und Bewertung der Planungsinstrumente	23
Tabelle 6	Aufbau Registerprotokoll	31
Tabelle 7	Unterschiede Dokumentationsmodelle	51
Tabelle 8	Bewertungskriterien Erfassungssysteme	59
Tabelle 9	Realisierungsmöglichkeiten Serversystem	66
Tabelle 10	Voraussetzungen zur Integration mit Versorgungssystemen	73
Tabelle 11	Semantische Zuordnung bei Transkription	74

Tabelle 12	Gegenüberstellung der Integrationsarten mit Forschungssystemen	84
Tabelle 13	Beispiel Ableitung eines Merkmals	93
Tabelle 14	Umsetzungsentscheidungen	96
Tabelle 15	Merkmalskatalog klinisches Myelomregister . . .	111
Tabelle 16	Plausibilitätsprüfungen klinisches Myelomregister	113
Tabelle 17	IT-Infrastruktur klinisches Myelomregister . . .	116
Tabelle 18	Test- und Implementierungsprotokoll	152
Tabelle 19	Testinstanz 1 für Test- und Implementierungs- protokoll	155
Tabelle 20	Testinstanz 2 für Test- und Implementierungs- protokoll	156

ABKÜRZUNGSVERZEICHNIS

CRF	Case Report Form
DBS	Datenbanksystem
EDC	Electronic Data Capture
EPA	Elektronische Patientenakte
EPL	Elektronische Patientenliste
ETL	Extract, Transform, Load
IDAT	Identifizierende Daten
IDMS	Identitätsdatenmanagementsystem
IR	Information Retrieval
KDAT	Klinische Daten
KIS	Krankenhausinformationssystem
MGUS	Monoklonale Gammopathie unbestimmter Signifikanz
MDMS	Metadatenmanagementsystem
PMS	Patientenmanagementsystem
RDMS	Registerdatenmanagementsystem
SDMS	Studiendatenmanagementsystem
VS	Virtueller Server

EINLEITUNG

GEGENSTAND UND BEDEUTUNG

Die Beobachtung und Dokumentation von Krankheits- und Therapieverläufen unter Realbedingungen hat in den letzten Jahren immer stärker an Bedeutung gewonnen und dazu beigetragen, dass klinische und wissenschaftliche Datensammlungen national und international unter dem Begriff des *klinischen Registers* aufgebaut und entwickelt wurden (Löbe et al. 2013; Pang et al. 2014; Plotnicki et al. 2013; Verband der forschenden Arzneimittelhersteller 2014; Viviani et al. 2014; Zylka-Menhorn 2015).

Als klinisches Register wird eine „aktive, standardisierte Dokumentation von Beobachtungseinheiten zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen“ bezeichnet (Müller et al. 2010). Das Ziel, die Behandlungs- und Versorgungsroutine abzubilden, verdeutlicht ihre große gesellschaftliche Bedeutung und hebt klinische Register von klinischen Studien ab. Während eine klinische Studie auf ein Patientenkollektiv mit klar definierten Ein- und Ausschlusskriterien und auf eine Hauptfragestellung hin optimiert ist, können in einem klinischen Register weitere Patienten aus der Versorgungsroutine eingeschlossen und therapeutische Langzeiteffekte, Nebenendpunkte und Begleiterscheinungen beobachtet werden (Rustenbach et al. 2011; Wegscheider 2004). Die Lücke zwischen interventionellen klinischen Studien einerseits und der Routinebehandlung unter Alltagsbedingungen andererseits lässt sich dadurch teilweise schließen (Marschner 2009). So kann ein klinisches Register ein bedeutendes Werkzeug zur Unterstützung klinischer Forschung sein, mit dessen Hilfe sich beispielsweise neue Hypothesen generieren oder Hinweise zur Fallzahlplanung ableiten lassen (Klinkhammer-Schalke et al. 2012). Der großen Bedeutung klinischer Register hat auch 2008 das Bundesministerium für Gesundheit Rechnung getragen und den flächendeckenden Ausbau der klinischen Krebsregistrierung zu einem vorrangigen Ziel des Nationalen Krebsplanes erklärt (Bundesministerium für Gesundheit 2008).

Die wachsende Anzahl methodischer Richtlinien, Vorgehensbeschreibungen und Leitlinien unterstreicht die Wichtigkeit klinischer Register und ihrer systematischen Planung und ermöglicht z. B. die schrittweise Definition von Zielen und reproduzierbare Ableitung von Merkmalen, Patientenkollektiven und Systemanforderungen (Gliklich u. Dreyer 2014; Müller et al. 2010).

PROBLEMATIK

Eine systematische Planung und ein standardisiertes Vorgehen sind eine wichtige Voraussetzung für den unmittelbaren und langfristigen Erfolg eines klinischen Registers.

Alle bekannten Lösungskonzepte setzen dabei implizit den Registeraufbau mit dem Aufbau eines neuen prolektiven wissenschaftlichen Dokumentationssystems gleich und berücksichtigen existierende wissenschaftliche Datenbestände und Infrastrukturen nur unzureichend, was wiederum die Praxistauglichkeit der Leitlinien einschränken kann. Dies kann dazu führen, dass neue Systeme parallel zu vorhandenen Systemen etabliert, Daten redundant und heterogen zu existierenden Datensammlungen erfasst und die verfügbaren Ressourcen ineffizient eingesetzt werden. Unmittelbare Auswirkungen auf Datenqualität und -nutzung sowie zusätzlicher finanzieller, personeller, organisatorischer und technologischer Aufwand können die Folge sein.

Klinische Register lassen sich nicht auf dieselbe Art und Weise und mit denselben Instrumenten planen und aufbauen wie klinische Studien. Sie erfordern ein spezifisches systematisches Vorgehen, eigene Methoden und Konzepte zur Integration in bestehende Informationssysteme einer klinisch-wissenschaftlichen Versorgungs- und Forschungseinrichtung. Leitfäden, die diese Aspekte auf Planungs- und Realisierungsebene umfassend berücksichtigen, fehlen bislang.

MOTIVATION

Die wissenschaftliche Dokumentation in einem klinischen Register basiert häufig auf den Daten zum Krankheits- und Therapieverlauf eines Patienten, die im Rahmen der Patientenversorgung erhoben und erfasst werden und im Krankenhausinformationssystem (KIS) verfügbar sind. Die Übertragung wissenschaftlich relevanter Merkmale aus den klinischen Informationssystemen in eine Datensammlung wird dabei meist händisch von medizinischen Dokumentaren durchgeführt. Dieser Vorgang, Informationen aus der Patientenversorgung manuell zu selektieren, wissenschaftlich nützlich aufzubereiten und zu übertragen, ist zeitlich aufwändig und fehleranfällig und wird durch monotone Routinearbeiten wie dem Anlegen neuer Patienten oder der Eingabe von Laborparametern zusätzlich erschwert. Die Integration eines klinischen Registers in die bestehenden Informationssysteme der Versorgung ist hilfreich, um neben einer manuellen Datenerfassung eine Sekundärnutzung klinischer Daten auch technisch zu ermöglichen.

Wissenschaftlich tätige Versorgungseinrichtungen, die über Jahrzehnte hinweg Patientenversorgung und klinische Forschung für eine Krankheitsentität praktizieren, verfügen oft über umfangreiche wissenschaftliche

Systeme und Datenbanken. Obwohl solche Datensammlungen einen Projektcharakter haben, häufig heterogen und eventuell redundant vorkommen, bilden sie das wissenschaftliche Fundament einer Forschungseinrichtung. Ein neues prolektives Dokumentationssystem geht aber immer einher mit einem Neubeginn und einer Neuerfassung. Durch die Integration von Altdaten und weiteren verfügbaren wissenschaftlichen Datensammlungen sind klinische Register in ihrem Nutzen deutlich erweitert, sowohl zu Beginn als auch in fortgeschrittenem Betriebsstadium bei der Beobachtung und Analyse von Therapie- und Langzeiteffekten.

Planung und Aufbau eines klinischen Registers sollten daher einhergehen mit der Integration in die existierenden Dokumentationssysteme der Versorgung zur Sekundärnutzung klinischer Daten sowie mit der Integration in existierende Datensammlungen der Forschung für quellübergreifende Auswertungen. Erst das Gesamtsystem aus prolektivem Dokumentationssystem und dessen Integration mit den Systemen der Versorgung und Forschung hat nachhaltigen und zukünftigen Nutzen für eine klinisch-wissenschaftliche Versorgungseinrichtung.

ZIELSETZUNG

Aus der Problematik lassen sich folgende Ziele ableiten, die im Rahmen dieser Dissertationsschrift erreicht werden sollen:

ZIEL 1 Entwicklung eines Leitfadens zur systematischen Planung eines klinischen Registers unter besonderer Berücksichtigung der Integration in existierende Informations- und Dokumentationssysteme von Versorgung und Forschung.

ZIEL 2 Anwendung des Leitfadens am Beispiel eines neu einzuführenden klinischen Myelomregisters am Universitätsklinikum Heidelberg.

AUFBAU DER DISSERTATIONSSCHRIFT

Diese Arbeit ist gemäß der zu erreichenden Zielen wie folgt aufgebaut:

Kapitel 2 enthält aktuelle Definitionen und Methoden zu klinischen Registern, systematischer Planung und beschreibt relevante Ansätze zur Integration von Daten und Informationen.

Kapitel 3 betrachtet vorhandene methodische Richtlinien und Leitlinien zu klinischen Registern im Detail, bewertet diese und leitet Anforderungen an ein Vorgehensmodell zur systematischen Planung eines klinischen Registers und an dessen IT-Infrastruktur ab.

Im Kapitel 4 wird ein Vorgehensmodell für die systematische Planung eines klinischen Registers entwickelt und darauf aufbauend eine Referenz-IT-Architektur entworfen, die besonders die Integration in existierende Informationssysteme von Versorgung und Forschung berücksichtigt. Vorgehensmodell und Referenz-IT-Architektur sind so beschrieben, dass sie als Leitfaden für neu einzuführende und aufzubauende klinische Register genutzt werden können.

Eine praktische Anwendung des zuvor entwickelten Leitfadens wird im Kapitel 5 am Beispiel des klinischen Myelomregisters des Universitätsklinikums Heidelberg dargestellt.

Die Arbeit schließt mit Diskussion und Ausblick in Kapitel 6.

Im folgenden Kapitel werden die Grundlagen, die für die Entwicklung eines Vorgehensmodells zur systematischen Planung eines klinischen Registers und einer geeigneten IT-Infrastruktur notwendig sind, beschrieben. Die Grundlagen enthalten aktuelle Definitionen und Methoden zu klinischen Registern und systematischer Planung sowie etablierte Ansätze zur Integration von Daten und Informationen.

2.1 KLINISCHE REGISTER

Der Begriff des *Registers* stammt aus dem lateinischen und bezeichnet sowohl die Methodik der Registrierung, d. h. das gezielte Sammeln von Informationen, als auch das dabei entstehende Produkt, die Datei oder Datensammlung selbst (Wegscheider 2004). Datensammlungen, die Gesundheitsinformationen enthalten, werden allgemein als *klinische Register* bezeichnet. In der Fachliteratur werden auch die Begriffe *Patientenregister*, *klinische Datenregister*, *Krankheitsregister* oder *Outcomeregister* verwendet.

Das Verständnis eines klinischen Registers als reine Datensammlung klinischer Informationen wurde 1974 von Brooke um Aspekte wie Systematik und Zweckmäßigkeit bei Datenerhebung und Zieldefinition erweitert (Brooke 1974). In vielen aktuellen Zeitschriftenartikeln und Definitionen ist dieser Schwerpunkt immer noch maßgeblich (Müller et al. 2010). Der größer werdenden Bedeutung klinischer Register wurden Gliklich und Dreyer gerecht, die keine reine Datensammlung oder Dokumentation definierten, sondern ein klinisches Register als „organized system“ bezeichneten, das deskriptive Methoden aus Beobachtungsstudien zur einheitlichen und zielgerichteten Datenerhebung und Auswertung nutzt (Gliklich u. Dreyer 2007).

Demnach bieten klinische Register das Potenzial den gesamten Krankheits- und Therapieverlauf eines Patienten von der Diagnose über die einzelnen Therapieabschnitte bis hin zu Nachsorge, Rezidiven, Überleben und Tod zu beobachten (Gliklich u. Dreyer 2014; Klinkhammer-Schalke et al. 2012). Diese personenbezogene Beobachtung und Datenerhebung erfolgt in der Versorgungsroutine und bildet den klinischen Zustand eines Patienten ohne zusätzliche, über die Routinebehandlung hinausgehende klinische Intervention ab (Aulitzky u. Becker 2011). Daher werden klinische Register häufig auch als *nicht-interventionelle Beobachtungsstudien* bezeichnet (Rustenbach et al. 2011).

Im Unterschied zu randomisierten, kontrollierten Interventionsstudien haben klinische Register in der Regel nur wenige Ein- und Ausschlusskriterien und erlauben die nicht-selektive Beobachtung von Patientengruppen, die aufgrund bestimmter Symptomkonstellationen wie Multimorbidität nicht an klinischen Studien teilnehmen dürfen (Wegscheider 2004). Das in klinischen Studien eingeschlossene Patientenkollektiv macht nur einen kleinen Teil der im Alltag zu versorgenden Patienten aus und führt häufig dazu, dass die therapeutische Wirksamkeit einer klinischen Intervention überschätzt wird (Rustenbach et al. 2011). Hinzu kommt, dass klinische Studien oft auf eine Hauptfragestellung hin optimiert sind und beendet werden, sobald der primäre Endpunkt erreicht ist (Wegscheider 2004). Aussagen über die therapeutische Alltagswirksamkeit eines in der Versorgungsroutine behandelnden Patienten sind damit oft nicht oder nur eingeschränkt möglich.

Mit Hilfe klinischer Register bietet sich die Möglichkeit, wissenschaftliche Fragen zu beantworten, die in klinischen Studien offen bleiben, aber weiterhin bedeutend für die ärztliche Praxis sind (Kathmann 2008; Wegscheider 2004). So lassen sich sowohl Langzeiteffekte und Nebenendpunkte abschätzen, also auch die Wechselwirkung von Therapiewahl, Wirksamkeit, Begleiterscheinungen und Krankheitsverläufen bestimmen und zur Behandlungsoptimierung nutzen (Klinkhammer-Schalke et al. 2012; Wegscheider 2004).

Klinische Register sind allerdings kein Konkurrenz- oder Ersatzkonzept zu klinischen Studien, sondern vielmehr eine weitere Methodik der klinischen Forschung, die die Lücke zwischen evidenzbasierter Forschung einerseits und der Alltagssituation andererseits schließen kann. So können Daten aus klinischen Registern beispielsweise die Konzeption von randomisierten, kontrollierten Interventionsstudien unterstützen, Hinweise für die Fallzahlplanung liefern oder Hypothesen zur Wirksamkeit von Behandlungen bei speziellen Patientensubgruppen generieren (DWG-Gesellschaft 2015; Klinkhammer-Schalke et al. 2012).

Allgemein lassen sich unter einem *guten klinischen Register* die Anforderungen zusammenfassen, Daten aus der Routinebehandlung zu erheben, um daraus Wirksamkeit, Effektivität und Sicherheit therapeutischer Maßnahmen zu evaluieren (Rustenbach et al. 2011). Dieser Anspruch, die Versorgungsroutine präzise zu beschreiben und gleichzeitig Dokumentation und Datenerhebung einheitlich, standardisiert und prolektiv zu realisieren, zeigt allerdings auch eine gewisse Widersprüchlichkeit: zwar soll die Datenerfassung denselben systematischen Kriterien klinischer Studien folgen, basiert aber letztendlich auf der nativen Beobachtung der Routinebehandlung, die naturgemäß nur wenig standardisiert ist (Rustenbach et al. 2011; Wegscheider 2004). Während sich die Datenerhebung in klinischen Studien durch Studienprotokolle, Erhebungsbögen, Therapie- und Visitenzeitpunkte kontrollieren lässt, sind in klinischen

Registern sowohl die individuellen Behandlungs- und Therapieverläufe, als auch die vorhandenen Datenquellen heterogener, unstrukturierter und unsystematischer. Klinische Register bilden *real world data* nicht nur ab, sondern basieren auch maßgeblich auf diesen und nutzen verstärkt Patientenakten, Arztbriefe oder Befunde der Routineversorgung anstatt spezifisch entwickelter und optimierter Erhebungsformulare (Gliklich u. Dreyer 2014; Trotter 2002).

2.2 SYSTEMATISCHE PLANUNG

Planung beschreibt die gedankliche Vorwegnahme des zukünftigen Handelns zur Erreichung der angestrebten Ziele mit den dafür zur Verfügung stehenden Mitteln (Wöhe u. Döring 2013). Im Gesundheitswesen, in dem die Rechte, die Sicherheit und das Wohlergehen der Patienten die höchsten anzunehmenden Güter darstellen, kommt auch der Planung medizinischer Dokumentationssysteme eine große gesellschaftliche Verantwortung zu (European Medicines Agency 1996). Medizinische Dokumentationssysteme, wie beispielsweise Patientenakten, Basis- und Tumordokumentationen, klinische Register, Studien und Archive weisen eine hohe Komplexität auf und erfordern ein methodisch angemessenes, umsichtiges und reproduzierbares Vorgehen (Leiner et al. 2011).

Sichergestellt werden diese Eigenschaften insbesondere dann, wenn die Planung *systematisch*, d. h. nach vorab formulierten Richtlinien und Handlungsanweisungen erfolgt (Leiner 1994). Ein solches Vorgehensmodell haben Leiner und Haux 1996 für die Planung medizinischer Dokumentationssysteme entwickelt (Leiner u. Haux 1996): in einem ersten Schritt werden zunächst die konkret zu lösenden Probleme so detailliert beschrieben, dass daraus die Dokumentationsziele abgeleitet werden können. Anschließend lassen sich aus jedem individuellen Ziel mehrere Fragen und Aufgaben bestimmen, aus deren Summe wiederum ein Entwurf des Dokumentationssystems mit Datenhaltung, Erfassung, IT- und Sicherungskonzept spezifiziert werden kann (Leiner et al. 2011).

Durch dieses mehrstufige Ableitungsschema, ausgehend von den Problemen über die Ziele hin zu den Anforderungen und dem Systementwurf, werden nur die Systemanforderungen hergeleitet, die sich eindeutig aus den Dokumentationszielen ergeben und die beim Aufbau des Dokumentationssystems mindestens zu berücksichtigen sind. Jede zusätzliche Anforderung führt zu zusätzlichem Aufwand und reduziert den Dokumentationsnutzen (Leiner et al. 2011). Dies kann sich in der Folge unmittelbar auf die Motivation zur Dokumentation auswirken, die weiter die Daten-, und diese die Ergebnisqualität beeinflussen kann. Entsprechend ist es notwendig, Anforderungen und Ziele, die durch den Systementwurf nicht oder nur mit unverhältnismäßig hohem Dokumentationsaufwand erreicht werden können, frühzeitig zu identifizieren, anzupassen und

dadurch Nutzen und Motivation zur Dokumentation zu verbessern (Leiner 1994).

Systematische Planung bedeutet aber auch, dass man sich zunächst auf abstraktem Niveau mit der Zielsetzung, den Anforderungen und den Aufgaben auseinandersetzen muss und konkret umzusetzende Punkte, wie beispielsweise der Merkmalskatalog erst im späteren Planungsprozess folgen. Wird die Planung hingegen zu früh zu spezifisch, kann dies dazu führen, dass die Menge an notwendigen Informationen überschätzt wird und mehr Daten erfasst werden, als für die Erreichung der Ziele eigentlich notwendig. Verstärkt wird diese Gefahr dadurch, dass in der medizinischen Forschung durch ein *Mehr* an Daten die Wahrscheinlichkeit steigt Korrelationen und Effekte nachzuweisen, und häufig ein zu hoher Detaillierungsgrad oder eine zu große Fallzahl angelegt wird (Gliklich u. Dreyer 2014). Tatendrang und Enthusiasmus einer schnellen Umsetzung können so die systematische Planung erschweren, oder ganz verdrängen (Viviani et al. 2014).

Für den Registeraufbau wird in mehreren Veröffentlichungen eine systematische Planung gefordert und auf das Vorgehensmodell von Leiner und Haux verwiesen (Gliklich u. Dreyer 2014; Müller et al. 2010), sodass dies im Kapitel 3 als existierende methodische Richtlinie beschrieben und den registerspezifischen Leitfäden gegenübergestellt wird.

2.3 INTEGRATION

Integration beschreibt die (Wieder-)Herstellung einer Einheit aus Teilen und wird in der Informatik als Sammelbegriff für verschiedene Vorgänge verwendet, bei denen mehrere Systeme, Komponenten oder deren Inhalte zu einem Gesamtsystem zusammengeführt werden. Die Form der Zusammenführung ist dabei abhängig von der Bezugsebene und kann in *Daten-, Funktions-, Prozess- und Präsentationsintegration* unterteilt werden (Huber 2014; Lehmann 2005). Im Rahmen dieser Dissertationsschrift wird insbesondere auf die *Datenintegration* eingegangen, mit dem Ziel redundante Erhebung und Speicherung von Daten zu vermeiden und eine Datenquelle für unterschiedliche Anwendungen und Ziele verfügbar zu machen. Eine Verallgemeinerung der Datenintegration ist die *Informationsintegration*, die die Zusammenführung von Daten innerhalb eines Kontextes und unter Wahrung ihrer Bedeutung vorsieht (Leser u. Naumann 2007). Der Begriff *Integration* wird, sofern nicht anders spezifiziert, im Folgenden synonym für *Informationsintegration* verwendet.

Integration ist immer eine Folge von Heterogenität, die wiederum das Ergebnis von Verteilung und Autonomie darstellt. Werden Informationssysteme physisch oder logisch in mehrere Einzelsysteme aufgeteilt, dann begünstigt die Autonomie der Systeme, die unabhängig voneinander konfiguriert, administriert und weiterentwickelt werden, unterschiedliche

Gestaltungsmöglichkeiten und Entscheidungen, die sich in heterogenen Systemen manifestieren (Herschel 2010). Aufgabe der Integration ist es Heterogenität auf unterschiedlichen Systemebenen zu überwinden. Eckstein differenziert dabei nach *technischer*, *syntaktischer* und *semantischer Heterogenität* (Eckstein 2011). Technische Heterogenität liegt beispielsweise dann vor, wenn sich Systeme in Hardware, Software oder Schnittstellen unterscheiden und verschiedene Abfragesprachen für Datenzugriff, Austausch und Kommunikation verwenden. Syntaktische Heterogenität bezeichnet unterschiedliche Darstellungen derselben Information durch verschiedene Codierungen, Datentypen, Formate, Modelle oder Schemata. Während sich technische und syntaktische Heterogenität in der Regel durch Standards, Transformations- und Abbildungsregeln überwinden lassen, sind semantische Konflikte nicht immer eindeutig zu identifizieren und aufzulösen. Erschwert wird dies häufig dadurch, dass keine Informationen zu Intension und Semantik einer Datensammlung oder eines Systems vorliegen, und indirekt über Bezeichnungen und Inhalte darauf geschlossen werden muss. Entsprechend sind nicht nur Daten Grundlage und Voraussetzung für Integration, sondern ebenso System-, Quellen- und Merkmalsbeschreibungen (Leser u. Naumann 2007). Typische Arten semantischer Heterogenität sind Benennungskonflikte mit synonymen oder homonymen Bezeichnungen, Wertebereichskonflikte, falls dieselben Merkmale unterschiedliche Ausprägungen haben, und Skalierungskonflikte bei verschiedenen, aber ineinander umrechenbaren Maßeinheiten (Maur u. Winter 2003). Semantische Heterogenität kann nur durch genaue Analyse und Kenntnis der Informationssysteme überwunden werden. Das semiotische Dreieck nach Ogden et al. veranschaulicht diesen Zusammenhang und beschreibt, dass eine Bezeichnung allein einen realen Gegenstand oder Sachverhalt nicht direkt, sondern nur über ein (gedankliches) Konzept eindeutig repräsentieren kann (Ogden et al. 1923). Zur Beschreibung von Konzepten und Beziehungen sind Wissensrepräsentationssprachen wie *Glossare*, *Taxonomien*, *Thesauri* oder *Ontologien* hilfreich (Adametz u. Billig 2009). Dabei werden Ansätze zur *Konzeptualisierung* und *Spezifikation* genutzt, um alle relevanten und gültigen Konzepte des Wirklichkeitsausschnitts festzulegen, zu benennen und die zugehörigen Eigenschaften und Integritätsbedingungen zu definieren. Das Ergebnis sind domänen- oder anwendungsspezifische *kontrollierte Vokabulare* zur eindeutigen Bezeichnung von Merkmalen und zur Erkennung semantischer Konflikte.

Da Heterogenität in der Praxis unvermeidbar ist, Systeme und Datensammlungen unterschiedlich gewachsen und etabliert sind, und trotzdem häufig der Wunsch nach einer einheitlichen, konsistenten und globalen Sicht auf die Daten besteht, sind verschiedene Methoden und Maßnahmen zur Sicherstellung von Interoperabilität und Überwindung technischer, syntaktischer und semantischer Heterogenität notwendig.

Standardisierungs- und Harmonisierungsbestrebungen können dabei initial unterstützen, eine effiziente und möglichst vollständige Zusammenführung verschiedener Datenstrukturen zu einer Einheit, ist allerdings in der Regel nur über einen mehrstufigen Integrationsprozess realisierbar. Die Ziele Einheitlichkeit, Vermeidung von Redundanzen und Mehrfach-erfassungen sind für klinisch-wissenschaftliche Versorgungseinrichtungen ebenso erstrebenswert, wie quellübergreifende Auswertungen und multidimensionale Analysen klinischer, administrativer und molekularbiologischer Gesundheitsdaten.

Der erste Schritt der Integration besteht üblicherweise in der Prüfung organisatorischer und datenschutzrechtlicher Rahmenbedingungen. Liegen Datenquellen außerhalb des eigenen Verantwortungsbereichs, wurden während des Behandlungskontextes erhoben oder schränken eine Sekundärdatennutzung ein, sind zusätzliche Ethikvoten, schriftliche Einverständniserklärungen und Datennutzungsvereinbarungen erforderlich (Pommerening u. Muscholl 2015). Sind diese Voraussetzungen erfüllt, dann gilt es, technische Heterogenität durch Schnittstellen, Nachrichtenaustauschformate und Übertragungsprotokolle zu überwinden, um Zugriff auf die gewünschten Datenquellen zu erlangen. Syntaktische Heterogenität wird nach Bedarf schrittweise gelöst und sieht die Transformation auf einheitliche Dateiformate (beispielsweise CSV oder XML), Strukturen (bspw. hierarchisch oder relational), Zeichencodierungen (bspw. UTF-8 oder ANSI) und Datenformate (bspw. Dezimalpunkt oder -komma) vor. Erfolgen Plausibilitäts- und Qualitätsprüfungen nicht in den Quellsystemen, sollte die Konsistenz der Datensammlungen während des Integrationsprozesses durch Prüfungen und Korrekturen unzulässiger, fehlender oder falscher Werte und dem Erkennen und Entfernung von Duplikaten sichergestellt werden. Integration wird durch die Kombination eines *Record Linkage* und einer *semantisch interoperablen Merkmalsfusion* realisiert (Leser u. Naumann 2007). Dabei werden Instanzen aus verschiedenen Datenquellen, die eindeutig miteinander verknüpft, und deren semantische Konflikte mit Hilfe kontrollierter Vokabulare gelöst werden, zu einem Datensatz zusammengeführt. Die Bereitstellung kann wahlweise in einem Dateiformat zur Auswertung mittels spezifischer Statistiksoftware oder in einem für Analysezwecke optimierten *Data Warehouse* erfolgen (Lusti 2002).

Zusätzliche Anforderungen und technische Komponenten, die bei Planung und Integration eines klinischen Registers in bestehende Informations- und Dokumentationssysteme von Versorgung und Forschung berücksichtigt werden, sind an den entsprechenden Stellen des Vorgehensmodells und der IT-Infrastruktur im Kapitel 4 mit direktem Bezug zu Konzeption und Realisierung enthalten.

Ein systematischer Prozess bei Planung, Aufbau und Betrieb eines klinischen Registers wird in mehreren nationalen und internationalen Veröffentlichungen gefordert und gilt als wichtiger Faktor für den unmittelbaren und langfristigen Erfolg eines Registers (Gliklich u. Dreyer 2014; Leiner et al. 2011; Mathis u. Wild 2008; Müller et al. 2010). Zwar wurde in den letzten Jahren vielfach über klinische Register, deren Besonderheiten und Gegensätzen zu klinischen Studien publiziert, doch wenige Arbeiten thematisieren Planung, Aufbau und Betrieb im Detail. Sie eignen sich nur eingeschränkt als Leitfaden oder Handlungsanweisung für ein eigenes Vorgehen. Daher wurde in einer Literaturrecherche explizit nach Leitfäden zu klinischen Registern gesucht und die identifizierten Veröffentlichungen hinsichtlich ihrer Eigenschaften und Anforderungen untersucht. Am Ende dieses Kapitels werden die existierenden Leitfäden gegenübergestellt und der Bedarf nach einem registerspezifischen Vorgehensmodell zur systematischen Planung und einer darauf aufbauenden IT-Infrastruktur abgeleitet.

3.1 LITERATURRECHERCHE

Die Literaturrecherche erfolgte als Kombination aus systematischer Recherche und Schneeballsystem (Brühl 2016). Dazu wurde in einem ersten Schritt eine Suchanfrage in der Literaturdatenbank *PubMed* durchgeführt und Leitfäden zu Registern bestimmt, die wissenschaftliche und forschungsnahe Zielsetzungen verfolgen. Da klinische Register häufig nationalen Regelungen wie Meldepflichten und Datenschutzgesetzen genügen müssen, wurde die Anfrage eingeschränkt auf Leitfäden, die in deutschen Versorgungseinrichtungen angewendet werden. Tabelle 1 zeigt die in *PubMed* formulierte Suchanfrage.

Tabelle 1: Suchanfrage in *PubMed* nach Handlungsanweisungen und Leitfäden zu klinischen Registern mit wissenschaftlichem Anwendungsgebiet und Nutzung in Deutschland.

```
registry[Title/Abstract] AND
research[Title/Abstract] AND
(guideline[Title/Abstract] OR guide[Title/Abstract]) AND
(germany[Title/Abstract] OR german[Title/Abstract])
```

Eine vollständige Prüfung der Ergebnisliste zeigte, dass nur die Veröffentlichung *Memorandum Register für die Versorgungsforschung* von Müller et al. die Anforderungen an einen umfassenden und vollständigen Leitfaden für ein klinisches Register erfüllt (Müller et al. 2010).

Anschließend wurde ein Schneeballsystem auf die im *Memorandum* referenzierten Literaturangaben angewandt und die Publikationen *Registries for Evaluating Patient Outcomes: A User's Guide* von Gliklich und Dreyer, *Systematic Planning of Clinical Documentation* von Leiner und Haux und *Register für klinische und gesundheitsökonomische Fragestellungen* von Mathis und Wild als weitere Leitfäden zu klinischen Registern identifiziert (Gliklich u. Dreyer 2010; Leiner u. Haux 1996; Mathis u. Wild 2008).

Während der detaillierten Durchsicht der Veröffentlichung von Mathis und Wild wurde festgestellt, dass die dargelegten methodischen und organisatorischen „Good Practice Strategien“ für klinische Register eine deutsche Übersetzung und Zusammenfassung des ausführlichen Leitfadens von Gliklich und Dreyer darstellen (Gliklich u. Dreyer 2007). Da damit kein weiterer Erkenntnisgewinn gegenüber dem Originalbeitrag hinsichtlich Planungsvorgehen und spezifizierten Anforderungen an ein klinisches Register einhergeht, wurde die Veröffentlichung aus nachfolgender Analyse ausgeschlossen. Zudem wurde die aktuelle Version des *User's Guide* verwendet, sodass neben den Leitfäden von Müller et al. und Leiner und Haux auch der Leitfaden von Gliklich und Dreyer im Folgenden im Detail betrachtet wird (Gliklich u. Dreyer 2014; Leiner u. Haux 1996; Müller et al. 2010).

Da in dieser Dissertationsschrift insbesondere auf systematische Planung, Aufbau und Betrieb eines klinischen Registers und auf die informationsverarbeitenden und speichernden Systeme und Komponenten Bezug genommen wird, werden die genannten Leitfäden auch nur hinsichtlich dieser Aspekte beschrieben und analysiert. Organisatorische, personelle und rechtliche Anforderungen, die in den Leitfäden teilweise umfangreich enthalten sind, werden hier nicht aufgeführt.

3.2 MEMORANDUM REGISTER FÜR DIE VERSORGUNGSFORSCHUNG

Das *Memorandum Register für die Versorgungsforschung* wurde 2010 vom *Deutschen Netzwerk Versorgungsforschung* als Bestandteil der zweiteiligen Reihe *Methoden für die Versorgungsforschung* in der Zeitschrift *Das Gesundheitswesen* veröffentlicht (Müller et al. 2010). Weitere Leitfäden der Reihe beschreiben Methoden zur gesundheitsökonomischen Evaluation, zur Erfassung der Lebensqualität und für die organisationsbezogene Versorgungsforschung. Das *Memorandum* für klinische Register wurde von 17 Fachexperten erstellt und wird von 45 nationalen Gesellschaften

getragen, unter anderem der *Deutschen Gesellschaft für Epidemiologie*, der *Deutsche Gesellschaft für Innere Medizin* und der *Technologie- und Methodenplattform für die vernetzte medizinische Forschung* (TMF).

Das Ziel ist die Bereitstellung eines Leitfadens zur Entwicklung und Qualitätsbeurteilung eines klinischen Registers. Der Umfang beläuft sich auf 16 Seiten und ist in folgende Kapitel unterteilt:

Tabelle 2: Gliederung *Memorandum Register für die Versorgungsforschung*.

-
1. Definition eines Registers
 2. Registerentwicklung
 3. Technische Organisation eines Registers
 4. Auswertung
 5. Berichterstattung
 6. Datenschutz, rechtliche und ethische Aspekte
 7. Kriterien zur Beurteilung der Qualität eines Registers
 8. Checkliste Registerqualität
-

In Kapitel 1 wird der Begriff des Registers definiert als „möglichst aktive, standardisierte Dokumentation von Beobachtungseinheiten zu vorab festgelegten, aber im Zeitverlauf erweiterbaren Fragestellungen, für die ein präziser Bezug zur Zielpopulation transparent darstellbar ist“. Dabei wird unter *aktiv* eine prolektive¹, von registereigenem Personal durchgeführte Datenerhebung, gemäß den jeweils formulierten Registerzielen, verstanden. Das *Memorandum* unterscheidet zwischen Krankheitsregistern, Produktregistern, Qualitätsregistern und Expositionsregistern zur Unterstützung klinischer Forschung, Monitoring der Patientensicherheit und Wirksamkeitsevaluierung in der Versorgungsroutine.

Die Registerentwicklung wird in die Phasen Planung, Aufbau und Umsetzung differenziert und soll fortlaufend in Form eines *Registerprotokolls*, entsprechend dem *Dokumentationsprotokoll* von Leiner und Haux festgehalten und spezifiziert werden (Leiner u. Haux 1996). Die Planungsphase beginnt mit dem Zweck und dem konkreten Einsatzgebiet des Registers und erfordert eine explizite Formulierung und Begründung des erwarteten Nutzens. Analog zu Leiner und Haux werden aus den Zielen eindeutige Fragestellungen und daraus die zu erhebenden Merkmale abgeleitet, die in Ziel-, Einfluss- und Störgröße unterteilt werden. Weiter sollen Ziel- und Quellpopulation über Ein- und Ausschlusskriterien definiert werden sowie im Hinblick auf die Rahmenbedingungen, ein oder mehrere Auswertungskollektive. Bei den Erhebungsverfahren ist die Art und Anzahl der teilnehmenden Zentren, der Meldeweg (papierbasiert

¹ In der Veröffentlichung wird der Begriff *prospektiv* zur Definition verwendet. Es ist allerdings davon auszugehen, dass *prolektiv* gemäß der Definition von Feinstein gemeint ist (Feinstein 1977).

oder elektronisch), der Zeitraum und die Organisation der Nachverfolgung zu regeln. Ausgehend von diesen Festlegungen lassen sich Datensatz, Datenerfassung, Datenmanagement, Analyseplan sowie rechtliche und ethischen Anforderungen weiter spezifizieren. Hierbei hervorzuheben ist die schrittweise Ableitung der Merkmale und die Definition eines Mindestdatensatzes, der nur obligatorische Merkmale enthält. Die Merkmale werden in einem Merkmalskatalog dargestellt und als logisches Datenmodell abgebildet, wobei nationale und internationale Datensätze zu berücksichtigen sind. Ausgehend von dem logischen Datenmodell und der Art der Meldung sind Erfassungsformulare zu entwerfen, Merkmalsdefinitionen und Kodierungsanleitungen festzulegen sowie Validierungsregeln, Auswertungsmethodik und Ergebnispräsentation anzugeben. Die Umsetzung sieht die Überführung des logischen Datenmodells in ein relationales Datenbankschema vor sowie Auswahl, Kauf und Parametrisierung der Hard- und Softwarekomponenten.

Die technische Organisation eines Registers wird im *Memorandum* in die vier Ebenen Geschäftsführung, IT-Management, Software und Hardware unterteilt. Die Geschäftsführung wird unabhängig von dem IT-Management betrachtet und übernimmt neben repräsentativen Aufgaben auch operative, wie die Erstellung und Pflege von Erfassungsformularen, die Qualitätssicherung oder das Fortschreiben des Registerprotokolls. Das IT-Management ist hingegen für das Vertragsmanagement, falls technische Komponenten an Drittanbieter ausgelagert sind, oder für die eigene Infrastruktur verantwortlich. Auf Softwareseite ist ein elektronisches Erfassungssystem (Electronic Data Capture; EDC) auszuwählen und zu parametrisieren sowie eine Webanwendung über eine verschlüsselte Client-Server-Infrastruktur zu realisieren. Physisch sollten mindestens zwei Server, idealerweise drei, zur Verfügung stehen, wobei einer für den Produktiveinsatz, einer zur Ausfallsicherung und ein weiterer als Test- und Entwicklungsumgebung genutzt werden soll. Virtuelle Server werden als Alternative zu physischen Server genannt. Weitere Server werden notwendig, falls datenschutzrechtliche Anforderungen eine Trennung von identifizierenden und klinischen Daten vorsehen. Dabei ist in Deutschland mindestens die Abstimmung mit der zuständigen Datenschutzkommission bzw. dem Landesdatenschutzverantwortlichen erforderlich und die Einholung eines positiven Ethikvotums wird empfohlen. Generell sind Klartextspeicherung und Pseudonymisierung immer zustimmungspflichtig und lediglich eine Vollanonymisierung der Daten ist ohne schriftliche Einverständniserklärung des Patienten möglich. Die Trennung identifizierender und klinischer Daten erfolgt auf unterschiedlichen und organisatorisch unabhängigen Systemen. Die Sicherung des Datenschutzes ist in einem Datenschutzkonzept schriftlich festzuhalten und Pseudonymisierungsverfahren, Datenübermittlung, Datenspeicherung und Zugriffsberechtigungen sind zu beschreiben.

Das *Memorandum* ist inhaltlich sehr umfassend, allerdings auch sehr stark darauf optimiert, alle wesentlichen Punkte im Zusammenhang mit klinischen Registern zu benennen. Eine detaillierte und anschauliche Beschreibung fehlt in den meisten Fällen und daher ist es für einen Anwender auch nicht immer möglich den nächsten Entwicklungsschritt systematisch nachzuvollziehen oder konkrete Handlungsanweisungen abzuleiten. Beispielsweise wird ein Registerprotokoll als erforderliches Planungsinstrument eingeführt, allerdings nicht auf dessen Inhalte eingegangen, sondern auf das *Dokumentationsprotokoll* von Leiner und Haux verwiesen. Dadurch entstehen aber im Detail Inkonsistenzen, die im *Dokumentationsprotokoll* nicht oder anders gelöst werden als vom *Memorandum* spezifiziert (näheres dazu, bei der Beschreibung des *Dokumentationsprotokolls* im Kapitel 3.4). Weiter geht das *Memorandum* bereits bei der Definition eines klinischen Registers von einem neuen und projektiven System aus und berücksichtigt vorhandene Datenquellen der Versorgung und Forschung nicht. Da klinische Register aber Instrumente der wissenschaftlichen Forschung zur personenbezogenen Beobachtung und Erfassung der Alltagssituation und der Routinebehandlung sind, basieren sie auch zum Großteil auf Informationen, die im Rahmen der Patientenversorgung entstehen und weniger auf spezifisch neu erhobenen Merkmalen. Entsprechend fehlt es an Konzepten zur Nutzung und Integration bestehender Informationen aus den Dokumentationssystemen der Versorgung, die frühzeitig und systematisch in der Planungsphase des klinischen Registers zu berücksichtigen sind.

3.3 REGISTRIES FOR EVALUATING PATIENT OUTCOMES: A USER'S GUIDE

Der Leitfaden *Registries for Evaluating Patient Outcomes: A User's Guide* ist ein im Rahmen des US-amerikanischen Gesundheitsprogramms *Effective Health Care Program* entstandener Bericht der *Agency for Healthcare Research and Quality* und wurde erstmals 2007 als Referenzleitfaden für den Aufbau, den Erhalt und die Auswertung klinischer Register veröffentlicht. Über einen Zeitraum von zwei Jahren konnten sich verschiedene Interessengruppen an der Entwicklung beteiligen und online verfügbare Entwurfsfassungen kommentieren. Für die zweite und dritte Auflage, 2010 und 2014, wurde das Handbuch von über 280 Autoren und Gutachtern überarbeitet und auf zwei Bände mit je über 300 Seiten aufgeteilt (Gliklich u. Dreyer 2014).

Das Ziel des *User's Guide* ist die Unterstützung von Aufbau, Implementierung, Analyse, Interpretation und Evaluation klinischer Register. Die Kapitelstruktur ist chronologisch entsprechend der Entwicklungsphasen eines Registers in folgende sechs Abschnitte gegliedert:

Tabelle 3: Gliederung *Registries for Evaluating Patient Outcomes: A User's Guide*.

Section I.	Creating Registries
Section II.	Legal and Ethical Considerations for Registries
Section III.	Operating Registries
Section IV.	Technical, Legal, and Analytic Considerations for Combining Registry Data With Other Data Sources
Section V.	Special Applications in Patient Registries
Section VI.	Evaluating Registries

Jedem Abschnitt vorangestellt ist eine Kurzfassung, die die wesentlichen Überlegungen einer Phase zusammenfasst und sich in mehrere Kapitel unterteilt. Beispielsweise werden in *Section I. Creating Registries* die Eigenschaften klinischer Register definiert und ausführlich Planung, Aufbau und Herleitung des Merkmalkatalogs beschrieben.

Der *User's Guide* verwendet die Bezeichnung *patient registry*, die als weitgehend synonym zu der im Rahmen dieser Dissertationsschrift genutzten Benennung *klinisches Register* angesehen werden kann.

Gliklich und Dreyer haben in der ersten Auflage des *User's Guide* ein Register als *organisiertes System* definiert, das beobachtende Studienmethoden nutzt, um den Verlauf einer Erkrankung in der Versorgungsroutine zu beschreiben (Gliklich u. Dreyer 2007). Dadurch gehen Register deutlich über reine Datensammlungen hinaus und machen, ähnlich wie bei klinischen Studien systematische Vorgehensbeschreibungen und methodische Leitlinien notwendig.

Zu den wesentlichen Aufgaben in der Planungsphase eines klinischen Registers gehört die Ausformulierung der Ziele, aus denen konkret zu beantwortende wissenschaftliche und klinische Fragen abgeleitet werden. Dabei sind Zweckmäßigkeit, Vollständigkeit und Machbarkeit zu berücksichtigen. Die Ziele sollten von einem Sponsor, d. h. dem Auftraggeber des Registers vorgegeben werden, ebenso wie Beobachtungsgegenstand und Zielpopulation. Darauf aufbauend sind die Interessengruppen, deren Zielvorstellungen und die Rahmenbedingungen zu bestimmen und gegebenenfalls die Fragestellungen entsprechend der gewünschten Genauigkeit der Ergebnisse anzupassen. Abhängig von der Größe des Registers, der Anzahl beteiligter Einrichtungen und den finanziellen Rahmenbedingungen ist ein Register-Team mit verschiedenen Experten(-gruppen) aufzubauen, Maßnahmen von Registerführung und Kontrollgremien sind zu definieren und ein Registerprotokoll sowie ein Projektplan ist zu entwickeln.

Beim Registeraufbau werden die in der Planungsphase bestimmten Anforderungen weiter spezifiziert. Dazu gehört die Umformung der

klinischen Fragen in quantifizierbare Merkmale, die Ermittlung relevanter Datenquellen, die Auswahl der Patienten und eine Rekrutierungsstrategie. Zusätzlich werden Elemente klinischer Beobachtungsstudien, wie die Wahl des Studiendesigns, die Definition einer Vergleichsgruppe, Fallzahlplanung und Randomisierung beschrieben. Damit ist der *User's Guide* sehr umfassend in seiner Betrachtung und deckt nicht nur Register ab, die klinische Verläufe darstellen, sondern auch Register, die exponierte mit nicht-exponierten Gruppen oder Erkrankte mit gesunden Probanden vergleichen. Entsprechend werden als Registerdesign die klassischen Studiendesigns von Kohorten- und Fall-Kontroll-Studien eingesetzt, einhergehend mit Fallzahlplanung und Randomisierung. Das Studiendesign wirkt sich zudem auf die zu erfassenden Merkmale aus, die sich beispielsweise bei Probanden und Patienten unterscheiden können. Die Wahl der Merkmale sollte geleitet sein von Sparsamkeit, Validität und der Fokussierung auf die Fragestellungen des Registers. In vielen Fällen ist es sinnvoll, obligatorische von optionalen Merkmalen zu unterscheiden und jeweils den Erhebungsaufwand, bestehende Standards und validierte Skalen zu berücksichtigen. Merkmale werden zusätzlich nach ihrer Datenquelle klassifiziert: *primären Datenquellen* wird dabei eine höhere Validität zugesprochen als sekundären Quellen, da Planung und Erhebung prolektiv erfolgen. Merkmale aus primären Datenquellen dienen unmittelbar der Beantwortung der formulierten wissenschaftlichen und klinischen Fragestellungen. Existieren Merkmale bereits in anderen, sogenannten *sekundären Datenquellen*, wie beispielsweise Patientenakten, administrativen Datenbanken oder Geburts- und Sterberegistern, so ist zu prüfen, ob eine Übernahme daraus möglich ist. Da der *User's Guide* die Annahme macht, dass sekundäre Datenquellen in einem elektronischen Format und in strukturierter Form vorliegen, wird nur der direkte Import betrachtet. Primäre Datenquellen werden im *User's Guide* als nachteilig hinsichtlich des Erhebungsaufwandes beschrieben, sekundäre Datenquellen hinsichtlich der Validität. Während die Herleitung der Merkmale gut und nachvollziehbar dargestellt ist, wird die Wahl des Erfassungssystems, in dem die Merkmale abgebildet und in dem die Datenerfassung erfolgen soll, nicht beschrieben.

Im *Abschnitt III. Operating Registries* werden typische Anforderungen an Datenerhebung und Qualitätssicherung, wie Dokumentation, Datenbereinigung, Überwachung, klinische Bewertung und Berichterstellung diskutiert. Die Beschreibung orientiert sich verstärkt an organisatorische Rahmenbedingungen und formuliert technische Eigenschaften elektronischer Erfassungssysteme nur ansatzweise. Der *User's Guide* unterscheidet zwar zweckmäßig zwischen klinischen Studien und Registern, behält diese Konsequenz allerdings nicht bei der Beschreibung des elektronischen Erfassungssystems bei und nennt keine typischen Funktionen und Charakteristiken von Registerdatenmanagementsystemen (RDMS), sondern

betrachtet Studiendatenmanagementsysteme (SDMS) als äquivalent und ausreichend. Da aber SDMS auf klinische Studien ausgerichtet sind, mit fest vorgegebenen Behandlungs- und Therapieschemata, eignen sich diese nur eingeschränkt, um patientenspezifisch individuelle Krankheits- und Therapieverläufe adäquat abzubilden. Eine genauere Differenzierung und klare Abgrenzung klinischer Register von klinischen Studien auf technischer Ebene bei den Erfassungssystemen ist daher sinnvoll.

Einen Großteil des Leitfadens machen rechtliche und ethische Aspekte aus, die bei Entwicklung und Betrieb eines klinischen Registers zu berücksichtigen sind. Da der *User's Guide* auf Initiative des US-Gesundheitsministeriums entwickelt wurde, ist er auf die amerikanische Gesetzgebung, insbesondere auf *Health Insurance Portability and Accountability Act (HIPAA)* ausgelegt und kann nur eingeschränkt auf nationale Register übertragen werden. Technische Lösungskonzepte, die deutschen Datenschutzerfordernungen genügen, sind daher notwendig und müssen bereits in der Planungsphase des klinischen Registers berücksichtigt werden.

Im Unterschied zum *Memorandum* setzt der *User's Guide* in seiner Definition und Beschreibung von klinischen Registern da an, wo das *Memorandum* endet: bei der Sekundärdatennutzung klinischer Informationen. Demnach ist ein klinisches Register kein rein prolektives System, sondern basiert auch auf Merkmalen, die primär für die Patientenversorgung erhoben und nur sekundär für das klinische Register genutzt werden. Allerdings fehlt es an dieser Stelle an konkreten Beschreibungen der verschiedenen technischen Möglichkeiten und notwendigen Komponenten, um vorhandene Informationen, die Bestandteil der Dokumentationssysteme der Versorgung sind, mit prolektiv erfassten Registerdaten zusammenzuführen. Damit wird auch eine informationstechnische Infrastruktur erforderlich, die über ein elektronisches Erfassungssystem hinausgeht, um Informationen geeignet zu extrahieren, zu transformieren, zu integrieren und den Benutzern anzuzeigen.

3.4 SYSTEMATIC PLANNING OF CLINICAL DOCUMENTATION

In klinischen Studien sind *Studienprotokolle* weit verbreitet und zählen zur guten klinischen Praxis (Good Clinical Practice; GCP). Ihr Aufbau und Inhalt wird im Detail in eigenen Leitlinien beschrieben und vorgegeben (International Conference on Harmonisation 1996; International Society for Pharmaceutical Engineering 2008). Auf dem Gebiet der medizinischen Dokumentationssysteme ist das sogenannte *Dokumentationsprotokoll* eine ähnliche Methodik und inhaltlich an die *Studienprotokolle* angelehnt (Leiner et al. 2011). Neben dem reinen Protokollierungsaspekt, der die reproduzierbare Ableitung von Entscheidungen begünstigt, dient das

Dokumentationsprotokoll insbesondere als standardisiertes Vorgehensmodell zur systematischen Planung von Dokumentationssystemen.

Das *Dokumentationsprotokoll* wurde von Leiner im Rahmen seiner Dissertation am *Institut für Medizinische Biometrie und Informatik* der *Universität Heidelberg* entwickelt (Leiner 1994) und von Leiner und Haux in der Zeitschrift *Methods of Information in Medicine* publiziert (Leiner u. Haux 1996).

Das *Dokumentationsprotokoll* basiert auf der Annahme, dass sich ein systematischer Planungsprozess in folgende vier Schritte unterteilt:

- Formulierung der Dokumentationsziele, abgeleitet aus konkret zu lösenden Problemen.
- Beschreibung der Aufgaben und Anforderungen an das zukünftige System, abgeleitet aus den einzelnen Dokumentationszielen.
- Spezifikation des Systementwurfs, abgeleitet aus der Gesamtheit an Aufgaben und Anforderungen.
- Festlegung technischer und organisatorischer Rahmenbedingungen für die Entwicklung und Einführung des Dokumentationssystems, abgeleitet aus dem Systementwurf.

Diese Schritte sind als eigene Abschnitte im *Dokumentationsprotokoll* enthalten, denen eine Einleitung, eine Beschreibung von Problemstellung und Zielsetzung sowie eine Bewertung der Ausgangsbasis vorangestellt ist. Der detaillierte Aufbau des Dokumentationsprotokolls entspricht der Gliederung in Tabelle 4.

Die Planung eines Dokumentationssystems erfolgt entsprechend der Gliederung und beginnt bei Protokollpunkt (o) *Einleitung*. Das Vorgehen ist dabei bewusst in umgekehrter Reihenfolge zum späteren Dokumentationsprozess gewählt, um frühzeitig eine Diskussionsgrundlage zu schaffen und die Systemeigenschaften vor der Realisierung nachvollziehbar abzuleiten und darzustellen. Beispielsweise fällt dies bei Protokollpunkt 3.5 auf, in dem Präsentationsformulare erstellt und mit fiktiven Daten Auswertungen und Darstellungsformen beschrieben werden, noch bevor die *Erhebungsformulare* (4.2.2), das logische *Dokumentationsschema* (4.1.1) und der *Merkmalskatalog* (4.1.2) vorliegen. Die Vorwegnahme der Ergebnisse und deren Darstellung ermöglicht frühzeitig die Bewertung und Reflektion der erwarteten und realisierbaren Dokumentationsziele und gegebenenfalls deren Anpassung. Weiter wird durch den mehrstufigen Ableitungsprozess *Probleme* → *Ziele* → *Aufgaben* → *Anforderungen* sichergestellt, dass die Mindestmenge an Systemanforderungen an ein Dokumentationssystem zuverlässig bestimmt und umgesetzt werden kann. Entsprechend lassen sich Anforderungen, die nicht unmittelbar aus den

Tabelle 4: Gliederung des *Dokumentationsprotokolls*.

-
- (0) Einleitung
 - (1) Problemstellung und Zielsetzung
 - (2) Bewertung der Ausgangsbasis
 - (3) Spezifikation der Dokumentationsaufgaben
 - (3.1) Fragen und Aufgabenstellung
 - (3.2) Auswahl der Fragen
 - (3.3) Standardisierte Erhebungs- und Dokumentationsmethoden
 - (3.4) Statistische Auswertungsmethoden
 - (3.5) Präsentationsformulare
 - (4) Entwurf des Dokumentationssystems
 - (4.1) Datenhaltung
 - (4.1.1) Entwurf des Dokumentationsschemas
 - (4.1.2) Merkmalskatalog
 - (4.2) Aufzeichnung der Daten
 - (4.2.1) Organisation des Aufzeichnungsvorgangs
 - (4.2.2) Entwurf der Erhebungsformulare
 - (4.3) Rechnerunterstützte Kommunikation
 - (4.4) Sicherungskonzept
 - (4.4.1) Datenschutz
 - (4.4.2) Datensicherung
 - (5) Rahmenbedingungen
 - (5.1) Projektorganisation
 - (5.1.1) Verantwortlichkeiten und Berechtigungen
 - (5.1.2) Zeitplanung
 - (5.2) Dokumentationswerkzeuge
 - (6) Änderungen gegenüber der letzten Protokollversion
-

zu lösenden Problemen und zu erreichenden Zielen hervorgehen, oder deren Realisierung weitreichende personelle, technische und zeitliche Auswirkungen haben, bereits in der Planungsphase erkennen und gegebenenfalls anpassen. Entscheidungen, die während Planung und Aufbau zu treffen sind, werden in dem *Dokumentationsprotokoll* schriftlich festgehalten und bleiben dadurch reproduzier- und begründbar. Ebenso wird eine vollständige und übersichtliche Darstellung aller Entscheidungen und ihrer Auswirkungen auf die Systemrealisierung unterstützt. Wird das *Dokumentationsprotokoll* auch in der Betriebsphase regelmäßig fortgeschrieben, gibt es zu jedem Zeitpunkt den aktuell gültigen Stand des Dokumentationssystems wieder und kann als vertragliche Grundlage eines Entwicklungsauftrages und als Systembeschreibung genutzt werden.

Das *Dokumentationsprotokoll* besitzt seine Vorteile darin, ein systematisches und konkretes Ableitungsschema vorzugeben, das von Anwendern direkt genutzt, umgesetzt und linear abgearbeitet werden kann. Allerdings fehlen relevante Angaben bezüglich der organisatorischen und technischen Realisierungsaspekte, der zu treffenden Entscheidungen und möglichen Alternativen. Zudem ist es auf klinische Dokumentationssysteme im Allgemeinen und weniger auf klinische Register im Speziellen ausgerichtet, sodass für einen registerspezifischen Einsatz Anpassungen,

wie die Integrationsmethodik oder die Organisation der Nachverfolgung (Follow-Up), notwendig werden.

Das *Dokumentationsprotokoll* basiert auf der Planung und dem Aufbau eines medizinischen Dokumentationssystems als neues prolektives System. Erhebungsformulare werden entsprechend den Dokumentationszielen und den zu lösenden Problemen entworfen und genutzt und Merkmale rein prolektiv, d. h. „nach einem vorher festgelegten Plan“ erfasst (Feinstein 1977). Dieses Vorgehen kann daher auch als *problemorientiert* bezeichnet werden, da die Systemanforderungen aus konkret zu lösenden Problemen abgeleitet werden. Bei klinischen Registern ist allerdings die Annahme, dass ausschließlich prolektive Daten erhoben werden, unzureichend und *retrolektive Daten*, also solche „die irgend jemand schon ermittelt hat, bevor der klinische Untersucher auf die Idee kam, diese Daten zu sammeln“ (Lorenz 1981) müssen zusätzlich berücksichtigt werden. Während der *User's Guide* sekundäre Datenquellen der Routinebehandlung und damit retrolektive Daten als wichtig für klinische Register einstuft, liegt der Schwerpunkt des *Dokumentationsprotokolls* auf prolektiven Daten, sodass ein gemeinsames Vorgehensmodell, dass Systemanforderungen systematisch und kombiniert herleitet, fehlt. Ein vollständig *datenorientiertes* Vorgehen, wie es häufig bei der Analyse großer Datenmengen (*BigData*) diskutiert wird, bei dem Ziele allein aus vorhandenen retrolektiven Datenquellen abgeleitet werden, ist davon allerdings explizit abzugrenzen. Stattdessen wird das problemorientierte Vorgehen auch weiter als grundlegend für ein klinisches Register erachtet, erfordert aber eine Rückkopplung und die Möglichkeit zur iterativen Überarbeitung der Dokumentationsziele auf Basis vorhandener retrolektiver Datenquellen der Versorgung und Forschung.

3.5 VERGLEICH UND BEWERTUNG

Die Durchsicht und Analyse der identifizierten Leitfäden zeigte, wie wichtig ein systematisches Vorgehen bei Planung, Aufbau und Betrieb eines klinischen Registers ist, da alle drei dies explizit darlegen und wiederholt betonen. Aus Anwendersicht ist allerdings festzuhalten, dass nicht jeder Leitfaden uneingeschränkt hierfür geeignet ist, und sich Ansatz, Zielgruppe und Einsatzzweck unterscheiden.

Das *Dokumentationsprotokoll* von Leiner und Haux bietet als einziges ein konkretes, systematisches Ableitungsschema, das allerdings nicht für klinische Register optimiert ist und keine inhaltlichen, organisatorischen und technischen Entscheidungsmöglichkeiten aufzeigt (Leiner u. Haux 1996). Das *Memorandum* von Müller et al. hingegen ist auf klinische Register in deutschen Versorgungseinrichtungen ausgelegt, enthält aber keine direkten Handlungsanweisungen oder umsetzbare Beschreibungen (Müller et al. 2010). Zudem betrachtet es, wie auch das *Dokumentations-*

protokoll, nur die prolektiven Aspekte eines Systems und geht nicht auf vorhandene Datenquellen der Versorgung ein, die für klinische Register allerdings ebenfalls relevant sind. Lediglich der *User's Guide* von Gliklich und Dreyer beschreibt retrolektive Daten und deren Systeme, nennt aber keine technischen Möglichkeiten zur Nutzung und Integration (Gliklich u. Dreyer 2014). Entsprechend wenig betrachten alle vorhandenen Leitfäden technische Komponenten, registerspezifische Erfassungssysteme und infrastrukturelle Aspekte. In Tabelle 5 werden ergänzende Eigenschaften der drei untersuchten Leitfäden mit direkter Gegenüberstellung aufgeführt.

Tabelle 5: Bewertung und Gegenüberstellung der Leitfäden *Memorandum Register für die Versorgungsforschung* von Müller et al., *Registries for Evaluating Patient Outcomes: A User's Guide* von Gliklich und Dreyer und *Systematic Planning of Clinical Documentation* von Leiner und Haux (Gliklich u. Dreyer 2014; Leiner u. Haux 1996; Müller et al. 2010).

	<i>MEMORANDUM</i>	<i>USER'S GUIDE</i>	<i>DOKUMENTATIONS-PROTOKOLL</i>
Eigene Zielsetzung	Leitfaden für Planung und Aufbau klinischer Register	Leitfaden für Planung, Aufbau und Betrieb klinischer Register	Schema zur systematischen Planung klinischer Dokumentationssysteme
Beschriebener Ansatz	<i>problemorientiert und prolektiv</i> Aufbau eines neuen Dokumentationssystems und Ableitung der Systemanforderungen aus zu lösenden Problemen und Zielen	<i>problemorientiert, prolektiv und retrolektiv</i> Aufbau eines neuen Dokumentationssystems und Ableitung der Systemanforderungen aus zu lösenden Problemen und Zielen; maximaler Ansatz, deckt umfassend verschiedene Szenarien ab	<i>problemorientiert und prolektiv</i> Aufbau eines neuen Dokumentationssystems und Ableitung der Systemanforderungen aus zu lösenden Problemen und Zielen
Anwenderzielgruppe	Projektleiter während Planungsphase; auf deutsches Gesundheitssystem ausgerichtet	Projektleiter großer / nationaler klinischer Register während Planungs-, Aufbau- und Betriebsphase; Extraktion der Kernaussagen für Fachexperten erforderlich; auf amerikanisches Gesundheitssystem ausgerichtet	Projektleiter vor Planungsbeginn; Fachexperten (u. a. IT, Statistik) während Planungsphase; international anwendbar
Allgemeine Eignung für klinische Register	<i>gut</i> als Einstiegs- und Übersichtsdokument vorwiegend für Planungs- und Aufbauphase nutzbar; Betriebsphase nur minimal beschrieben	<i>sehr gut</i> als Nachschlagewerk für Planungs-, Aufbau- und Betriebsphase nutzbar	<i>gering</i> für Planungs-, Aufbau- und Betriebsphase nutzbar; aber nicht speziell für klinische Register ausgelegt
Eignung für systematische Planung klinischer Register	<i>gering</i> konkretes Vorgehensmodell muss selbstständig aus Beschreibung abgeleitet werden	<i>gering</i> konkretes Vorgehensmodell muss selbstständig aus Beschreibung abgeleitet werden	<i>sehr gut</i> gibt praktikables Vorgehensmodell für systematischen Planung vor

Tabelle 5: (Fortsetzung)

	<i>MEMORANDUM</i>	<i>USER'S GUIDE</i>	<i>DOKUMENTATIONS- PROTOKOLL</i>
Eignung für den Aufbau klinischer Register	<i>sehr gering</i> eigene inhaltliche Systemanforderungen nicht ableitbar; konkrete technische Realisierung nicht ableitbar	<i>gering</i> eigene inhaltliche Systemanforderungen nur mühevoll ableitbar, aufgrund umfangreicher Beschreibung vieler verschiedener Szenarien und Rahmenbedingungen; konkrete technische Realisierung nicht ableitbar	<i>gering</i> eigene inhaltliche Systemanforderungen lassen sich ableiten; konkrete technische Realisierung aber nicht ableitbar
Beschreibt informationstechnische Realisierung und Systemkomponenten für klinische Register	<i>sehr gering</i> benennt Systemkomponenten auf organisatorischer Ebene; technische Beschreibung gering; nennt keine Werkzeuge	<i>sehr gering</i> beschreibt Systemkomponenten und Integration auf organisatorischer Ebene; keine technische Beschreibung	<i>nicht vorhanden</i>
Unterscheidet klinische Register von klinischen Studien	<i>teilweise</i> beschreibt Unterschiede, geht aber nicht auf Spezifika klinischer Register ein; insb. nicht bei der technischen Beschreibung der Systemkomponenten; betrachtet Register als <i>light Version</i> einer klinischen Studie	<i>teilweise</i> beschreibt Unterschiede, geht aber nicht auf Spezifika klinischer Register ein; betrachtet Register als <i>light Version</i> einer klinischen Studie	<i>nein</i>
Umfang	<i>teilweise zu gering</i> 12 Seiten; bestimmte Aspekte werden nur angerissen (z. B. Planungsphase), andere sehr ausführlich beschrieben (z. B. Auswertung)	<i>sehr hoch</i> ca. 600 Seiten aufgeteilt auf zwei Bände	<i>angemessen</i> 10 bzw. 20 Seiten inklusive Beispiele in Publikation bzw. Dissertation

Tabelle 5: (Fortsetzung)

	<i>MEMORANDUM</i>	<i>USER'S GUIDE</i>	<i>DOKUMENTATIONS- PROTOKOLL</i>
Detailierungsgrad	<i>gering</i> benennt alle relevanten Aspekte, beschreibt aber nur wenige ausreichend	<i>sehr hoch</i> beschreibt verschiedene Rahmenbedingungen und die resultierenden Konsequenzen für Planung und Aufbau sehr ausführlich	<i>angemessen</i> benennt und erläutert die Gliederungspunkte zweckmäßig
Nachvollziehbarkeit	<i>gut nach</i> Entwicklungsphasen gegliedert; Phasen aber weitestgehend unabhängig voneinander; schnelle Einarbeitung möglich	<i>gering</i> insbesondere durch Zusammenfassungen gegeben; nach Entwicklungsphasen gegliedert; Phasen aber verschachtelt; lange Einarbeitung notwendig	<i>sehr gut nach</i> Entwicklungsphasen gegliedert; Phasen aufeinander aufbauend; schnelle Einarbeitung möglich
Wissenschaftlichkeit	<i>hoch</i> Publikation in eigener nationaler Fachzeitschrift; 17 Autoren; unterstützt von 45 nationalen Mitgliedergesellschaften	<i>sehr hoch</i> Publikation als eigenständiges Dokument; mehrjähriger und mehrstufiger Review-Prozess; 286 Autoren und Reviewer	<i>hoch</i> Dissertation und Peer-Review; Publikation in anerkannter internationaler Fachzeitschrift; 2 Autoren
Aktualität	<i>gering</i> Erst- und einmalige Veröffentlichung 2010; keine weitere Neuauflage	<i>hoch</i> Erstveröffentlichung 2007; bisher drei Neuauflagen; zuletzt 2014	<i>angemessen</i> Erstveröffentlichung 1994; regelmäßige Publikation als Teil eines Lehrbuchs; zuletzt 2012; Neuauflagen nicht notwendig
Planungsbeispiele	<i>nicht vorhanden</i>	<i>mehrere vorhanden</i> reale Planungs- und Umsetzungsbeispiele; keine technische Umsetzung enthalten	<i>vorhanden</i> fiktives Planungsbeispiel
Basiert auf	geringfügig an <i>Dokumentationsprotokoll</i> angelehnt; inhaltlich stark an <i>User's Guide</i> angelehnt; referenziert auf weiterführende Literatur	an keinen bekannten Leitfaden angelehnt; referenziert umfangreich auf internationale Publikationen und Standards	Inhaltlich und strukturell an Studienprotokolle angelehnt

3.6 ABGELEITETE ANFORDERUNGEN AN EINEN LEITFADEN

Die Anforderungen an ein Vorgehensmodell zur systematischen Planung eines klinischen Registers und an dessen IT-Infrastruktur werden im Folgenden aufgeführt und begründet. Die Ableitung der einzelnen Punkte erfolgte auf Grundlage der betrachteten Leitfäden sowie weiterer Literatur, die zusätzliche Eigenschaften, Anforderungen und Anregungen enthielt.

PLANUNG NACH EINEM DEFINIERTEN VORGEHENSMODELL UND ABLEITUNGSSCHEMA Ähnlich zu klinischen Studien und medizinischen Dokumentationssystemen sollte einem klinischen Register ein Protokoll zugrunde liegen, das die Vorgehensweise festlegt und wissenschaftliche Systematik und Angemessenheit belegt (Rustenbach et al. 2011). Dieses Vorgehensmodell ist vorab zu definieren und sollte Systemanforderungen systematisch aus konkret zu lösenden Problemen und zu erreichenden Zielen ableiten (Leiner et al. 2011).

PLANUNG UNTER BERÜCKSICHTIGUNG RETROLEKTIVER SYSTEME UND DATENQUELLEN Systematische Planung resultiert in der Regel in prolektiven Daten, die allerdings bei klinischen Registern nur einen Teil der Gesamtdatenmenge ausmachen (Gliklich u. Dreyer 2014). Notwendig ist daher der Umgang mit Systemen und Datenbeständen, die unabhängig vom klinischen Register entstanden sind und primär anderen Fragestellungen und Einsatzzwecken dienen. Hierbei sollte zwischen Systemen der Versorgung und Forschung und deren Eigenschaften hinsichtlich Standardisierung und Zugriffsmöglichkeiten differenziert werden. Ein ausschließlich prolektives Vorgehen birgt die Gefahr, dass das Register isoliert von bestehenden Systemen und Infrastrukturen aufgesetzt wird und dadurch in seinem Nutzen eingeschränkt bleibt.

GEMEINSAME HERLEITUNG DER REGISTERZIELE AUS PROBLEM- UND DATENORIENTIERTEM VORGEHEN Die systematische Planung eines klinischen Registers sollte nicht nur nach dem problemorientierten Vorgehen des *Dokumentationsprotokolls* von Leiner und Haux erfolgen, sondern auch das datenorientierte Vorgehen *in Ansätzen* berücksichtigen. Dabei sind frühzeitig in der Registerplanung die vorhandenen retrolektive Datenquellen der Versorgung und Forschung zu beachten, zu analysieren und mit den definierten Registerzielen iterativ abzugleichen.

ABBILDUNG DER VERSORGUNGSREALITÄT UND INDIVIDUELLER KRANKHEITS-, BEHANDLUNGS- UND THERAPIEVERLÄUFE
Klinische Register bilden die Versorgungsroutine von Patienten

ab und erfordern daher ein elektronisches System zur Erfassung und Verarbeitung individueller Krankheits-, Behandlungs- und Therapieverläufe (Gliklich u. Dreyer 2014). Studiendatenmanagementsysteme sind hierfür nur eingeschränkt geeignet, da das Dokumentationsmodell klinischer Studien fest vorgegeben ist, klinische Register allerdings variabel kombinier- und wiederholbare Abschnitte, Therapie- und Behandlungsphasen benötigen.

HOHE FLEXIBILITÄT BEI ÄNDERUNG DER REGISTERZIELE, FRAGESTELLUNGEN UND RAHMENBEDINGUNGEN Klinische Register sind häufig auf lange Laufzeiten ausgelegt, wodurch sich im Verlauf Beobachtungsschwerpunkte verschieben oder neue Forschungsfragen ergeben (Müller et al. 2010). Zusätzlich können neue Datenquellen erschlossen werden und sich Registerziele verändern. Dies alles erfordert eine hohe Flexibilität, Erweiterbarkeit und Adaptationsfähigkeit auf technischer, organisatorischer und inhaltlicher Ebene, die bereits in der Planungsphase und im Vorgehensmodell zu berücksichtigen sind.

NUTZUNG VON DATENQUELLEN AUS DER VERSORGUNGSROUTINE FÜR WISSENSCHAFTLICHE FRAGESTELLUNGEN Die Datenquellen der Versorgungsroutine sind im Unterschied zu Registerdaten in der Regel nur wenig standardisiert. Zur Sekundärdatennutzung klinischer Informationen für wissenschaftliche Fragestellungen sind daher, in Abhängigkeit der System- und Quelleigenschaften, verschiedene Verfahren erforderlich zur Extraktion, Transformation und Integration in das Erfassungssystem des Registers.

INTEGRATION IN BESTEHENDE DOKUMENTATIONS- UND INFORMATIONSSYSTEME VON VERSORGUNG UND FORSCHUNG Sollen neben Datenquellen der Versorgung auch retrolektive Daten wissenschaftlicher Systeme für gemeinsame wissenschaftliche Auswertungen genutzt werden, dann sind unabhängige Anwendungssysteme dafür nicht ausreichend und neben einem elektronischen Erfassungssystem ist eine IT-Infrastruktur aufzubauen, die die Integration pro- und retrolektiver Daten ermöglicht.

Diese Anforderungen wurden im Kapitel 4 in dem Vorgehensmodell für die systematische Planung eines klinischen Registers und in der Referenz-IT-Architektur berücksichtigt und umgesetzt.

In diesem Kapitel wird ein Vorgehensmodell für die systematische Planung und den Aufbau eines klinischen Registers entwickelt. Das Vorgehensmodell ist so aufgebaut, dass Systemanforderungen aus den Registerzielen einheitlich und nachvollziehbar abgeleitet und gleichzeitig alle relevanten Planungs- und Realisierungsaspekte systematisch dokumentiert werden können. Neu gegenüber dem *Dokumentationsprotokoll* von Leiner und Haux ist die Berücksichtigung registerspezifischer Eigenschaften, retroaktiver Daten und die Integration in die existierenden Systeme einer klinisch-wissenschaftlichen Versorgungseinrichtung (Leiner u. Haux 1996).

Aufbauend auf dem Vorgehensmodell wird eine Referenz-IT-Architektur entworfen. Diese umfasst verschiedene Anwendungssysteme sowie Kommunikations- und Integrationsaspekte, die zum Aufbau und Betrieb eines integrierten klinischen Registers erforderlich sind. Die Referenz-IT-Architektur wird so beschrieben, dass Hinweise zur Entscheidungsunterstützung bei der Auswahl und der Entwicklung der Anwendungssysteme aufgeführt werden.

Gemeinsam können das Vorgehensmodell und die Referenz-IT-Architektur als Leitfaden für Planung, Aufbau und Betrieb verschiedener klinischer Register und deren Integration in die existierenden Informationssysteme von Versorgung und Forschung genutzt werden.

4.1 VORGEHENSMODELL ZUR SYSTEMATISCHEN PLANUNG

Angelehnt an das *Dokumentationsprotokoll* wurde ein Vorgehensmodell entwickelt, das als Referenz für die systematische Planung eines klinischen Registers dienen kann. Dieses Vorgehensmodell, im Folgenden als Registerprotokoll bezeichnet, macht einen Gliederungsvorschlag zur angemessenen und reproduzierbaren Beschreibung und Herleitung der Ziele und Eigenschaften eines klinischen Registers und seiner Integration.

Das *Dokumentationsprotokoll* verfolgt einen problemorientierten Ansatz, bei dem die Anforderungen an ein Dokumentationssystem aus konkret zu lösenden Problemen abgeleitet werden. Bei klinischen Registern ist dies oft allein nicht ausreichend und erfordert ein erweitertes Vorgehen, das zusätzlich zum problemorientierten Ansatz auch die vorhandenen Datenquellen der Versorgung und Forschung mitberücksichtigt. Während prolektive Daten primärer Datenquellen das unmittelbare Ergebnis des problemorientierten Vorgehens sind, werden im Registerprotokoll

auch vorhandene retrolektive klinische Informationen sekundärer Datenquellen einbezogen.

Das Registerprotokoll gliedert sich in sieben Abschnitte und erstreckt sich über die drei Entwicklungsphasen Planung, Aufbau und Betrieb. Die Planungsphase, als zentrale Phase in der Registerentwicklung teilt sich dabei auf in die beiden Protokollpunkte 2 Definition der Registerziele und 3 Bewertung der Ausgangssituation. Die Spezifikation der Planung und der Aufbau des Registersystems folgen in Protokollpunkten 4 und 5. In Protokollpunkt 6 werden die für den Registerbetrieb notwendigen personellen, zeitlichen und technischen Ressourcen beschrieben und in Protokollpunkt 7 die Änderungen gegenüber der letzten Protokollversion. Tabelle 6 zeigt den gesamten Aufbau des Registerprotokolls im Überblick.

4.1.1 *Beschreibung der Inhalte*

Im Folgenden werden die Inhalte der einzelnen Protokollpunkte des Registerprotokolls im Detail beschrieben.

1 Einleitung

Es sind die am klinischen Register beteiligten Einrichtungen, Personen und Interessengruppen zu charakterisieren sowie der Gegenstand des Registers festzulegen und die Motivation für den Aufbau zu begründen.

1.1 Beteiligte Einrichtung(en)

Die am klinischen Register beteiligten Einrichtungen werden kurz beschrieben. Dabei sind nicht nur die Versorgungseinrichtungen selbst, sondern alle weiteren beteiligten Institute, Abteilungen usw. zu nennen. Eine Einrichtung gilt dann als beteiligt, wenn sie klinische Registerdaten selbst erhebt, erfasst oder zur Verfügung stellt.

Die Beschreibung sollte knapp sein, aber mindestens die Einrichtungen benennen und ihre vorgesehenen Aufgaben charakterisieren.

1.2 Beteiligte Interessengruppen und Personen

Es sind die Interessengruppen und Personen zu identifizieren, die an Planung, Aufbau, Betrieb und Auswertung des klinischen Registers beteiligt sind. Da an einem klinischen Register unterschiedliche Gruppen und Personen mitwirken, können diese auch auf unterschiedliche Weise von einem klinischen Register profitieren. Dadurch, dass die Interessengruppen frühzeitig definiert werden, kann zielgerichtet deren Motivation bestimmt und berücksichtigt werden. Dies kann Auswirkungen auf die Registerziele (siehe Protokollpunkt 2.2), die Nutzung der Daten (4.1), die Datenerhebung (4.3.2) und die Datenerfassung (5.2.1) haben.

Tabelle 6: Aufbau des Registerprotokolls. Die Kennzeichnung hinter jedem Gliederungspunkt kodiert den Unterschied zum *Dokumentationsprotokoll* von Leiner und Haux (Leiner u. Haux 1996):

- * wurde geringfügig geändert
- ** wurde inhaltlich erweitert
- *** wurde neu eingefügt

1	Einleitung**
1.1	Beteiligte Einrichtung(en)**
1.2	Beteiligte Interessengruppen und Personen**
1.3	Gegenstand und Motivation*
2	Definition der Registerziele**
2.1	Problemstellung*
2.2	Zielsetzung**
3	Bewertung der Ausgangssituation**
3.1	Aktueller Kenntnisstand**
3.2	Vorhandene Datenquellen**
3.3	Infrastrukturelle Anbindung***
3.4	Möglichkeiten und Grenzen des Registers**
4	Spezifikation des Registers**
4.1	Fragestellungen*
4.2	Patientenkollektiv**
4.2.1	Zielkollektive*
4.2.2	Auswertungskollektive**
4.3	Registermethodik**
4.3.1	Behandlungs- und Versorgungsroutine***
4.3.2	Erhebungsmethodik**
4.3.3	Integrationsmethodik***
4.3.4	Standardisierte Erhebungs- und Dokumentationsmethodik*
4.3.5	Statistische Auswertungsmethodik**
4.3.6	Präsentationsmethodik und Formulare*
5	Entwurf des Registers**
5.1	Datenhaltung**
5.1.1	Entwurf des Registerschemas**
5.1.2	Merkmalskatalog**
5.2	Datenmanagement**
5.2.1	Organisation der Datenerfassung*
5.2.2	Organisation des Follow-Up***
5.2.3	Rechnerunterstützte Datenerfassung**
5.2.4	Rechnerunterstützte Kommunikation**
5.2.5	Sicherung der Datenqualität***
5.2.6	Entwurf der Erfassungsformulare*
5.3	Sicherungskonzept*
5.3.1	Datenschutz*
5.3.2	Datensicherung*
6	Betrieb des Registers**
6.1	Verantwortliche und Berechtigungen**
6.2	Zeitplanung*
6.3	Informationstechnische Werkzeuge**
7	Änderungen gegenüber der letzten Protokollversion*

Typische Interessengruppen in klinischen Registern sind die Registerleitung, wissenschaftlich tätige Ärzte, medizinische Dokumentare, administrative Verwaltung, informationstechnische Mitarbeiter und Patienten.

Optional kann zwischen *primären* und *sekundären Interessengruppen* unterschieden werden: primäre Interessengruppen nutzen die Registerdaten zur Beantwortung der Fragestellungen von Protokollpunkt 4.1 oder sind verantwortlich für Planung, Aufbau, Betrieb und Auswertung, sekundäre Interessengruppen profitieren von dem Wissen, sind aber nicht elementar für den Aufbau des Registers.

Bei der Zahl beteiligter Interessengruppen sollte stets auf Zweckmäßigkeit geachtet werden, da durch mehr beteiligte Interessengruppen und Personen auch die Liste der Registerziele und folglich der zu erfassenden Merkmale steigen kann.

1.3 Gegenstand und Motivation

Es wird die Motivation für den Registeraufbau, die klinische Problematik und der zu beobachtende Wirklichkeitsausschnitt beschrieben. Dabei ist der Handlungsbedarf explizit zu begründen, sodass in späteren Projektphasen geprüft werden kann, ob die Problematik weiterhin besteht, oder mit dem Register gelöst werden konnte. Optional kann die Motivation auch nach den in Protokollpunkt 1.2 identifizierten Interessengruppen gegliedert werden.

Zusätzlich ist zu begründen, warum ein klinisches Register gegenüber einer kontrollierten randomisierten Studie bevorzugt wird.

2 Definition der Registerziele

Die Ableitung der Registerziele erfolgt in einem ersten Schritt aus den konkret zu lösenden Problemen (Protokollpunkt 2.1). Sobald die vorhandenen Datenquellen in Protokollpunkt 3.2 spezifiziert wurden, werden die Registerziele erneut bewertet und gegebenenfalls iterativ überarbeitet.

2.1 Problemstellung

Die in Protokollpunkt 1.3 dargestellte klinische (Gesamt-)Problematik wird näher beschrieben und in Form einzelner, konkret zu lösender Probleme präzisiert. Je genauer ein Problem formuliert werden kann, desto einfacher und reproduzierbarer die Ableitung der zugehörigen Ziele.

2.2 Zielsetzung

Zu jedem in Protokollpunkt 2.1 formulierten Problem werden ein oder mehrere Ziele definiert, die mit Hilfe des klinischen Registers erreicht werden sollen. Dabei sollte auf Überprüfbarkeit und reproduzierbare Herleitung geachtet werden. Zudem sind *allumfassende* Register und

die gleichzeitige Betrachtung zu vieler Probleme, zu integrierender Datensätze und Ziele zu vermeiden.

3 Bewertung der Ausgangssituation

Es werden die medizinischen, organisatorischen und finanziellen Rahmenbedingungen beschrieben, die das klinische Register betreffen. Daraus sollen Bedarf, Notwendigkeit und Machbarkeit des Registers analysiert und begründet werden.

3.1 Aktuelle Versorgungssituation und Kenntnisstand

Es wird beschrieben, wie die Routineversorgung des zu beobachtenden Patientenkollektivs erfolgt, welche Schwerpunkte bei Diagnostik und Therapie bestehen und welche Behandlungsoptionen existieren. Diese Darstellung gibt den Beobachtungsgegenstand des klinischen Registers wieder. Dabei sollte allen in Protokollpunkt 1.1 genannten Versorgungseinrichtungen mindestens eine Aufgabe in dem Krankheits-, Therapie- und Nachbeobachtungsverlauf zugeordnet werden.

Weiter sollte der aktuelle Wissensstand bezüglich Erkrankung und Therapie aus klinischen Publikationen ermittelt und wiedergegeben werden.

Die Einordnung der aktuellen Versorgungssituation und des klinischen Kenntnisstandes in die eigene Zielsetzung aus Protokollpunkt 2.2 ist zweckmäßig, um den Handlungsbedarf des klinischen Registers zu belegen.

3.2 Vorhandene Datenquellen

Es sind die vorhandenen Datenquellen zu identifizieren und zu beschreiben, die in das klinische Register integriert werden sollen. Anschließend sind die in Protokollpunkt 2.2 definierten Ziele zu prüfen und festzustellen, ob Merkmale in Form sekundärer Datenquellen bereits vorhanden sind, oder ob diese zunächst erhoben werden müssen. In vielen Fällen erfordert dies eine iterative Überarbeitung von Problemstellung, vorhandenen Datenquellen und Zielsetzung. Beispielsweise kann diese Betrachtung ergeben, dass ein formuliertes Ziel nicht, oder nur mit sehr viel Erhebungsaufwand zu erreichen ist, oder weitere Zielsetzungen unter Verwendung einer Datenquelle möglich sind, die zuvor unberücksichtigt blieben.

Hilfreich für diese Entscheidung kann folgendes Ableitungsschema sein:

1. Feststellen, ob für eine Zielsetzung die erwarteten Merkmale bereits in sekundären Datenquellen existieren.
2. Falls ja, feststellen, ob die Daten bzw. die Datenquellen zugänglich sind.

3. Falls ja, feststellen, ob die Daten in der notwendigen Qualität vorliegen.

Falls mindestens eine Frage mit *nein* beantwortet wird, ist eine Erhebung und prolektive Erfassung des Merkmals erforderlich.

Ein Ergebnis dieser Betrachtung kann auch sein, dass sich alle Probleme und die daraus abgeleiteten Zielsetzungen mit den Merkmalen aus den vorhandenen Datenquellen beantworten lassen. In so einem Fall ist eine redundante Datenerhebung und -haltung zu vermeiden und das Hauptaugenmerk auf Datenintegration zu legen.

Bei der Beschreibung der sekundären Datenquellen sollte ein qualitativer Unterschied zwischen klinischen *Routinedaten* und *Forschungsdaten* gemacht werden: Daten, die den Krankheits- und Therapieverlauf eines Patienten beschreiben, stammen meist aus einem Informationssystem der Versorgung, beispielsweise einem Patientenmanagementsystem (PMS) oder einem System zur Verwaltung elektronischer Patientenakten (EPA). Die Integration erfolgt in der Regel über eine sogenannte *Transkription*. Bei der Integration von Forschungsdaten, abgespeichert in wissenschaftlichen Datensammlungen wie Spreadsheets oder Datenbanken, ist zu klären, ob das klinische Register eine Forschungsdatenquelle permanent ablösen soll, in diesem Falle ist eine *Datenmigration* erforderlich, oder ob beide Systeme koexistieren und gemeinsam für wissenschaftliche Auswertungen genutzt werden können, realisierbar über eine *Verlinkung*. In beiden Fällen sollte zusätzlich zu der Beschreibung der Datensätze auch die Identifikation der Eigentümer und der Verfahrensweisen für Migration und Verlinkung erfolgen. Sind Forschungsdatenquellen nicht verfügbar oder zugänglich, erfordert dies eine erneute Transkription aus den Systemen der Versorgung oder eine wiederholte Datenerhebung. Beides sollte im Sinne der multiplen Verwendung möglichst vermieden werden.

Eine Einteilung der sekundären Datenquellen nach der Art der geplanten Datenintegration in Transkription, Migration und Verlinkung ist zweckmäßig.

3.3 Infrastrukturelle Anbindung

Ist ein klinisches Register in eine wissenschaftliche Versorgungseinrichtung eingebettet, wie ein Universitätsklinikum oder ein Forschungsinstitut, dann existieren neben den behandelnden Kliniken auch zentrale, administrative und medizinteoretische Einrichtungen. Diese können personelle, räumliche und logistische Ressourcen frei oder in Form von Dienstleitungen zur Verfügung stellen.

Zur Sicherung von Nachhaltigkeit und Vermeidung einer *Insellösung* ist es sinnvoll, frühzeitig die infrastrukturelle Anbindung und die angebotenen Dienstleistungen zu prüfen und das dort vorhandene Expertenwissen bereits bei der Planung des klinischen Registers zu

nutzen. Dadurch kann vermieden werden, dass Systeme redundant aufgebaut, Prozesse wiederholt entwickelt und zusätzliches Personal eingestellt oder geschult werden muss.

3.4 Möglichkeiten und Grenzen des Registers

Die Machbarkeit des Registers ist realistisch einzuschätzen. Dazu gehört neben der Betrachtung der Zielsetzung auch die Berücksichtigung des erwarteten Umfangs, charakterisiert durch die Anzahl an erwarteten Merkmalen, beteiligten Personen und einzuschließenden Patienten. Zusätzlich sind Dauer und Aufwand der Patientennachverfolgung zu beachten und die Nachhaltigkeit der Finanzierung zu klären.

4 Spezifikation des Registers

Ausgehend von den Zielen des klinischen Registers in Protokollpunkt 2.2 und unter Berücksichtigung der medizinischen, organisatorischen und finanziellen Rahmenbedingungen in Protokollpunkt 3 werden hier die detaillierten (wissenschaftlichen) Fragestellungen (Protokollpunkt 4.1), das Patientenkollektiv (4.2) und die gesamte Registermethodik (4.3), bestehend aus Behandlungs-, Erhebungs-, Integrations-, Erfassungs-, Auswertungs- und Präsentationmethodik, spezifiziert. In der Summe werden dadurch die Anforderungen festgelegt, denen das klinische Register genügen muss, um die oben definierten Zielsetzungen zu erreichen.

4.1 Fragestellungen

Zu jedem in Protokollpunkt 2.2 aufgeführten Ziel werden eine oder mehrere (wissenschaftliche) Fragen abgeleitet. Die Fragen sollten dabei möglichst eindeutig und präzise formuliert werden. Da aus einer Fragestellung unmittelbar ein oder mehrere Merkmale in Protokollpunkt 5.1 spezifiziert werden, ist die Ableitung der *richtigen* Fragen von zentraler Bedeutung. Daher sollten auch mehrere Experten, unter anderem *Mediziner*, *Datenmanager* und *Statistiker* miteinbezogen werden. Es ist denkbar, in einem ersten Schritt möglichst viele Fragen zu spezifizieren und anschließend eine Auswahl unter Berücksichtigung der wissenschaftlichen Relevanz und des erwarteten Dokumentationsaufwands zu treffen. Um diesen Prozess transparenter zu gestalten, besteht auch die Möglichkeit, zunächst nach obligatorischen und optionalen Fragestellungen zu unterscheiden.

Für Ziele aus Protokollpunkt 2.2, die mit Hilfe vorhandener Datenquellen erreicht werden können, sind die abgeleiteten Fragestellungen entsprechend zu kennzeichnen, um im weiteren Verlauf der Spezifikation des Registers den Verweis auf Datenintegration, anstelle auf Datenerhebung, nachvollziehbar festzuhalten.

4.2 Patientenkollektiv

Zu den in Protokollpunkt 4.1 formulierten Fragestellungen werden die Patientenkollektive definiert. Dabei wird zwischen Zielkollektiv (Protokollpunkt 4.2.1) und Auswertungskollektiv (4.2.2) unterschieden. Diese Unterscheidung ist abhängig davon, ob die Schlussfolgerungen des klinischen Registers nur für das beobachtete Patientenkollektiv gelten, in dem Fall ist nur ein Auswertungskollektiv zu definieren, oder ob die Aussagen auch für eine Grundgesamtheit verallgemeinert werden können.

Da abhängig von einer bestimmten Fragestellung das Auswertungs- und/oder das Zielkollektiv variieren kann, sind gegebenenfalls mehrere Kollektive zu beschreiben. Zusätzlich sollte jede Fragestellung in Protokollpunkt 4.1 auf das entsprechende Kollektiv verweisen.

4.2.1 Zielkollektive

Es sind die Patientenkollektive zu definieren, für die die Ergebnisse des klinischen Registers verallgemeinert gelten. Dies erfordert eine möglichst explizite Beschreibung der im Register zu beobachtenden Patientengruppe. Entsprechend werden Ein- und Ausschlusskriterien festgelegt, die verglichen zu klinischen Interventionsstudien allerdings weniger restriktiv sein sollten.

4.2.2 Auswertungskollektive

Es werden die Patientenkollektive definiert, die tatsächlich in dem klinischen Register eingeschlossen und ausgewertet werden können. Neben den in Protokollpunkt 4.2.1 beschriebenen Ein- und Ausschlusskriterien sind Auswertungskollektive in der Regel weiter eingeschränkt, unter anderem durch geographische und demographische Faktoren, oder durch nicht vorliegende Patienteneinverständnisse. Ein Zielkollektiv muss durch das entsprechende Auswertungskollektiv geeignet repräsentiert sein.

4.3 Registermethodik

Es werden die verschiedenen Aufgaben und Verfahrensweisen des klinischen Registers beschrieben, die sich allgemein unterteilen lassen in medizinische Behandlung (Protokollpunkt 4.3.1), Erhebung (4.3.2) oder Integration (4.3.3), Erfassung (4.3.4), statistische Auswertung (4.3.5) und Ergebnispräsentation (4.3.6).

4.3.1 Behandlungs- und Versorgungsmethodik

Die Behandlung der Patientenkollektive wird beschrieben und die relevanten Versorgungszeitpunkte wie beispielsweise Diagnose, Therapie und Nachbeobachtung werden identifiziert. Dabei sollten auch die durchgeführten Maßnahmen, wie beispielsweise Laboruntersuchungen,

molekulare oder genomische Analysen, aufgeführt werden. Hilfreich ist es zudem, den Ort der Versorgung anzugeben, um zwischen zentralisierter und verteilter Behandlung, die mehrere niedergelassene Ärzte umfasst, zu unterscheiden.

Das Ergebnis ist ein klinischer Untersuchungsplan, der den Versorgungsablauf in der Routine, die Zeiträume und die durchgeführten Verfahren beschreibt.

4.3.2 Erhebungsmethodik

Für Fragestellungen, die eine Erhebung prolektiver Daten erfordern und sich nicht mit Hilfe vorhandener Datensätze beantworten lassen, sind aus den Versorgungszeitpunkten des vorherigen Protokollpunktes die Erhebungszeitpunkte der Merkmale abzuleiten. Zusätzlich sind die entsprechenden Erhebungseinheiten, meist Patienten oder Ärzte sowie die Erhebungsinstrumente, wie Interviews, Fragebögen oder Tagebücher, festzulegen. Eine Feststellung von Merkmalsherkunft, Zuverlässigkeit und erwartetem Erhebungsaufwand erleichtert eine spätere Bewertung der Datenqualität. Beispielsweise lassen sich Gesundheitsfaktoren oder Symptome direkt von Patienten erheben, sind allerdings rein subjektiv. Ärzte hingegen können sehr gut klinische Eindrücke, Diagnosen oder Krankheitsstadien erheben.

4.3.3 Integrationsmethodik

Sind Ziele und die entsprechenden Fragestellungen in Protokollpunkt 4.1 mit Merkmalen aus vorhandenen Datenquellen zu beantworten, dann wird an dieser Stelle die Art der Datenintegration (Transkription, Migration, Verlinkung), der Zeitpunkt, die Häufigkeit und die Umsetzung pro Fragestellung und zugeordneter Datenquelle festgelegt.

4.3.4 Standardisierte Erfassungs- und Dokumentationsmethodik

Es wird für alle Fragestellungen, unabhängig von Erhebung oder Integration geprüft, ob eine standardisierte Dokumentation der Merkmale möglich ist, beispielsweise ob anerkannte, nationale oder internationale Datensatz- oder Merkmalsdefinitionen existieren. Die Nutzung standardisierter Datensätze ist wichtig, da sie Interoperabilität und Datenaustausch zwischen verschiedenen Systemen ermöglicht und zusätzlich eine validierte Merkmalsdefinition sicherstellt. Lassen sich keine entsprechenden Beschreibungen identifizieren, sollte geprüft werden, ob standardisierte und akzeptierte Ordnungssysteme, wie beispielsweise *ICD*, *OPS* oder *SNOMED*, genutzt werden können.

4.3.5 Statistische Auswertungsmethodik

Die verschiedenen geplanten statistischen Auswertungs- und Analysemethoden werden pro Fragestellung genannt und gegebenenfalls im Detail beschrieben. Dabei sind auch die Zeitpunkte der Auswertung und der Zweck, beispielsweise interne Qualitätssicherung oder Ergebnispräsentation, festzulegen.

Klinische Register können mitunter sehr komplexe Datenstrukturen haben, aufgrund wiederholter prolektiver Datenerfassung und multi-zentrischer Datenerhebung, wodurch eine frühe Einbeziehung einer statistischen Beratungsstelle zweckmäßig ist.

4.3.6 Präsentationsmethodik und Formulare

Für jede geplante Auswertung, die in Protokollpunkt 4.3.5 spezifiziert wurde und die in regelmäßigen Abständen wiederholt durchgeführt werden soll, wird ein Präsentationsformular entworfen. Darin wird die Art der Darstellung als Bericht oder Präsentation und die Form der erwarteten Ergebnisse festgelegt, beispielsweise wie numerische Werte deskriptiv oder grafisch dargestellt werden sollen.

Die Spezifikation von Präsentationsformularen mit fiktiven Werten illustriert die erwarteten Ergebnisse und deren Repräsentation. Missverständnissen bei den definierten Fragestellungen, den Erhebungsmethoden, den integrierten Merkmalen und den statistischen Auswertungsmethoden, lässt sich damit frühzeitig begegnen.

5 Entwurf des Registers

Die in Protokollpunkt 4 spezifizierten Planungsaspekte werden für Datenhaltung (Protokollpunkt 5.1), Datenmanagement (5.2) und Sicherungskonzept (5.3) weiter konkretisiert.

5.1 Datenhaltung

Es wird ein Datenmodell entworfen, das die zu erfassenden Informationseinheiten und deren Eigenschaften beschreibt und zueinander in Beziehung setzt. Das (semantische) Datenmodell dient als Abbildung des Wirklichkeitsausschnitts, der durch das klinische Register beobachtet und durch die Fragestellungen festgelegt wird, und kann in ein physisches Datenmodell überführt werden.

5.1.1 Entwurf des Registerschemas

In einem ersten Schritt wird festgelegt, welche Domänen im klinischen Register zu erfassen sind. Eine Domäne entspricht dabei einer Menge mehrerer zu dokumentierender Informationseinheiten (Instanzen) mit gleichen Eigenschaften. Typische Domänen in klinischen Registern sind *Patient*, *Diagnose*, *Therapie*, *Medikation* usw. Abgeleitet werden Domänen aus den in Protokollpunkt 4.2 angegebenen Patientenkollektiven und

den in 4.3.1 spezifizierten Versorgungszeitpunkten der Behandlungsmethodik. Jede Domäne besitzt Eigenschaften / Merkmale, die im nächsten Protokollpunkt festgelegt werden und die zur Beantwortung der Fragestellungen (Protokollpunkt 4.1) notwendig sind.

Anschließend werden die identifizierten Domänen zueinander in Beziehung gesetzt und Abhängigkeiten sowie der Zeitbezug festgelegt. Auch ist es notwendig zu beschreiben, ob Domänen wiederholbar sind oder variabel zu unterschiedlichen Zeitpunkten in der Versorgungsroutine auftreten können. Maßgeblich hierfür sind Behandlungs- (4.3.1) und Erhebungsmethodik (4.3.2).

Sind Daten aus vorhandenen Datenquellen zu integrieren, müssen deren semantische und physische Datenmodelle analysiert und auf die neu spezifizierten Domänen abgebildet werden.

5.1.2 Merkmalskatalog

Ausgehend von den Fragestellungen in Protokollpunkt 4.1 werden die Merkmale abgeleitet, die zur Beantwortung der Fragen notwendig sind. Dabei sollte zwischen neu zu erhebenden und zu integrierenden Merkmalen unterschieden werden. Dies ist hilfreich, da vorhandene Datenquellen bereits Merkmalseigenschaften enthalten und vorgeben, die hier wiederverwendet werden können. Dabei ist allerdings darauf zu achten, dass die Merkmalseigenschaften in der existierenden Datenquelle den erwarteten / benötigten Merkmalseigenschaften entsprechen oder dass dies durch eine Transformation sichergestellt werden kann.

Für jedes Merkmal sollte mindestens *Variablenname*, *Bezeichnung* auf dem Erhebungsformular, *Datentyp* und *Wertemenge* festgelegt werden. Hilfreich ist zudem die Angabe des *Quelldokuments* und einer formalen *Definition*. Englische Variablenamen sind deutschen vorzuziehen. Mehrsprachige Bezeichnungen sind in deutschsprachigen Registern optional.

Ist für ein Merkmal eine Wertemenge notwendig, wie bei einer Aufzählung, dann bietet es sich an, die Ausprägungen durch eine numerische Codierung weiter zu strukturieren. Dabei sollten über mehrere Merkmale hinweg die Codierungen konsistent beibehalten werden, sodass beispielsweise die Ausprägung *nein* immer mit *0* codiert wird.

Sind Datenquellen mit Registerdaten zu verlinken, dann müssen zusätzliche Merkmale für die eindeutige Identifikation der Beobachtungseinheiten spezifiziert werden.

5.2 Datenmanagement

Es sind die organisatorischen und technischen Prozesse zu entwerfen, die für die Erfassung der klinischen Registerdaten relevant sind.

5.2.1 Organisation der Datenerfassung

Für jedes in Protokollpunkt 5.1.2 spezifizierte Merkmal wird der Dokumentationsvorgang beschrieben. Dieser ist unter anderem von der erwarteten Aktualität und Zeitnähe zwischen Erfassung und Erhebung sowie von der Dauer und Art der Meldung (papierbasiert oder elektronisch) abhängig. Ein Merkmal, das ein medizinischer Dokumentar aus einem externen, papierbasierten Arztbrief in das klinische Register übertragen soll, erfordert beispielsweise ein anderes personelles und technisches Vorgehen als die Integration strukturierter Laborparameter aus dem KIS.

5.2.2 Organisation des Follow-Up

Es werden Zeitraum und Durchführung der Nachverfolgung der Patientenkollektive festgelegt. Dafür sind zunächst die Merkmale, die zur Charakterisierung des Langzeitverlaufs notwendig sind, zu identifizieren. Weiter ist es wichtig zu bestimmen, wie der Meldevorgang geregelt sein soll. Hinweise darauf gibt die in Protokollpunkt 4.3.1 beschriebene Behandlungs- und Versorgungsmethodik. Bei regelmäßigen Aufenthalten der Patienten in den beteiligten Versorgungseinrichtungen kann die Meldung durch den behandelnden Arzt erfolgen, andernfalls besteht die Möglichkeit die Patienten oder ihre niedergelassenen (Haus-)Ärzte postalisch anzuschreiben und den Gesundheitsstatus zu erfragen.

Auch wenn die Nachverfolgung ein Teil der Datenerfassung ist, sind meist eigene Prozesse zur Realisierung notwendig und daher die Beschreibung der Durchführung in einem eigenen Protokollpunkt zweckmäßig.

5.2.3 Rechnerunterstützte Datenerfassung

Aus den bisherigen Planungen ergeben sich Anforderungen an ein elektronisches Datenerfassungssystem, die hier spezifiziert werden. Prinzipiell besteht bei klinischen Registersystemen immer die Möglichkeit ein eigenes System zu entwickeln, oder ein bestehendes System auszuwählen und zu parametrisieren. Abhängig von der infrastrukturellen Anbindung in Protokollpunkt 3.3 können auch bereits Systeme vorhanden sein, die auf Eignung hin zu prüfen sind.

Standardeigenschaften eines Registersystems sind Erfassungformulare zur Dateneingabe und Datenbanken zur Speicherung. Zum Stand der Technik gehören Client-Server-Architekturen und der Zugriff auf das Registersystem über einen Webbrowser. Zudem sind mehrere Server zur Ausfallsicherung und zur Trennung klinischer von identifizierenden Daten sinnvoll.

5.2.4 Rechnerunterstützte Kommunikation

Der Datenaustausch und die Kommunikation mit anderen Anwendungssystemen innerhalb der beteiligten Versorgungseinrichtungen ist notwendig zur Integration vorhandener Datenquellen. Für Merkmale, die gemäß Protokollpunkt 5.2.2 nicht manuell erfasst, sondern elektronisch übertragen werden sollen, sind Kommunikationsschnittstellen zu entwerfen, die technische und strukturelle Eigenschaften definieren. Aufgrund der infrastrukturellen Anbindung des klinischen Registers sind häufig bereits Systeme, wie beispielsweise ein PMS oder ein EPA-System vorhanden und damit auch die entsprechenden Systemschnittstellen. Diese sind zu analysieren und das Registersystem ist so anzupassen, dass eine Kommunikation möglich wird. Unter anderem zu definieren ist das Dateiformat, die Syntax, die erwarteten Informationseinheiten (Instanzen) und die Merkmalsmenge, die notwendig ist um Daten aus dem Registersystem in ein anderes Anwendungssystem zu übertragen und umgekehrt.

5.2.5 Sicherung der Datenqualität

Für die einzelnen zu erfassenden Merkmale sind Maßnahmen zur Sicherung der Datenqualität zu entwerfen. Dazu gehören unter anderem Plausibilitätsprüfungen und logische Beziehungen. Zweckmäßig ist eine Unterscheidung in *harte* und *weiche Regeln*: harte Regeln werden aktiv, wenn Format oder Datentyp verletzt werden, weiche Regeln bei unplausiblen Werten.

Zusätzlich ist festzulegen, wie fehlende Merkmale zu handhaben sind, beispielsweise wenn eine Untersuchung nicht durchgeführt wurde, oder Werte nicht verfügbar sind. Diese sogenannten *NULL-Werte* sollten explizit als solche gekennzeichnet und die entsprechenden Merkmale nicht einfach leer gelassen werden. Bei Auswahllisten sind zusätzliche Optionen wie *nicht dokumentierbar* oder *unbekannt* zweckmäßig. Bei numerischen Werten ist ein einheitliches Vorgehen festzulegen, beispielsweise eine negative Codierung oder ein zusätzliches Merkmal, dass ein Wert nicht vorhanden ist.

Weitere Qualitätsindikatoren der medizinischen Forschung sind in der Literatur beschrieben, unter anderen in Nonnemacher et al. (Nonnemacher et al. 2007).

5.2.6 Entwurf der Erfassungsformulare

Es sind die Formulare der Merkmalerfassung zu entwerfen. Erfassungsformulare, oder auch Case Report Forms (CRFs), sind papierbasierte oder elektronische Bögen, die die Eingabe der in Protokollpunkt 5.1 spezifizierten Merkmale erlauben. Alle zuvor spezifizierten Merkmalseigenschaften wie *Bezeichnungen*, *Datentypen*, *Auswahloptionen*, *NULL-Werte*, *logische Beziehungen* und *Plausibilitätsprüfungen* werden

auf den Erfassungsformularen grafisch dargestellt. Sie bilden die Schnittstelle zum medizinischen Dokumentar und sind so zu entwerfen, dass ein manueller Dokumentationsvorgang möglichst gut abgebildet wird. Dabei bietet es sich an, für jede identifizierte Domäne mindestens ein eigenes Erfassungsformular zu entwerfen und die zugehörigen Merkmale in eine sinnvolle strukturelle Reihenfolge zu bringen.

5.3 Sicherungskonzept

Das Sicherungskonzept legt technische und organisatorische Anforderungen zum Schutz der Daten vor Datenmissbrauch (Protokollpunkt 5.3.1) und Verlust (5.3.2) fest.

5.3.1 Datenschutz

Es sind die technischen und organisatorischen Maßnahmen zum Schutz der (personenbezogenen) Registerdaten vor Datenmissbrauch zu beschreiben. Der Schutz des Rechts auf informationelle Selbstbestimmung, der Persönlichkeit und der Privatsphäre sind in Deutschland im Bundesdatenschutzgesetz (BDSG), in Landesdatenschutzgesetzen (LDSG) und, in einigen Bundesländern ergänzend, in Landeskrankenhausgesetzen (LKHG) geregelt. Dabei gilt, dass die Erhebung, Verarbeitung und Nutzung personenbezogener Daten grundsätzlich verboten ist, es sei denn, es liegt eine explizite Erlaubnis, in den meisten Fällen eine Einwilligung des Patienten, vor. Anonymisierte Daten, d. h. solche, die keinen direkten Rückschluss auf den Patienten ermöglichen und damit nicht mehr personenbeziehbar sind, bedürfen keiner expliziten Erlaubnis, sind aber für klinische Register meist nur von geringem praktischen Nutzen, da für Erhebung, Erfassung, Integration und Nachverfolgung immer personenbeziehbare Merkmale notwendig sind – sowohl organisatorisch als auch informationstechnisch. Pseudonymisierung und Klartextspeicherung hingegen erfordern das schriftliche Einverständnis des Patienten.

Pseudonymisierung wird in der Regel verwendet, um klinische von identifizierenden Daten zu trennen. Hierfür sind die Verfahren und Pseudonyme festzulegen und zu beschreiben.

Gegebenenfalls kann an dieser Stelle auch auf das Datenschutzkonzept verwiesen werden, in dem alle datenschutzrechtlichen Belange in Schriftform festgehalten sind und das dem Datenschutzbeauftragten der beteiligten Versorgungseinrichtungen zur Prüfung vorgelegt werden muss. Es ist zweckmäßig die zuständigen Datenschutzbeauftragten (und Ethikkommissionen) in einem frühen Entwicklungsstadium des klinischen Registers einzubeziehen, um Bedenken auszuräumen und Verfahrensweisen zu optimieren.

5.3.2 Datensicherung

Es sind die technischen und organisatorischen Maßnahmen zum Schutz der Registerdaten vor Datenverlust zu beschreiben. Beispielsweise erhöhen mehrere redundante Server die Ausfallsicherheit, parallel arbeitende Systeme verbessern die Verfügbarkeit und regelmäßige Systemaktualisierungen steigern die Zuverlässigkeit. Auf Ebene des Registersystems sind zudem Vorkehrungen zu treffen, um gespeicherte Daten vor (un-)beabsichtigter Manipulation und Löschung zu sichern. Dazu gehört unter anderem eine detaillierte Beschreibung der Zugangsvergabe, der Zugriffsberechtigungen und die Protokollierung aller Nutzeraktivitäten und Veränderungen.

6. Betrieb des Registers

Die personellen, zeitlichen und technischen Ressourcen, die dem klinischen Register während des Betriebs zur Verfügung stehen, werden beschrieben.

6.1 Verantwortliche und Berechtigungen

Für Aufbau und Betrieb eines klinischen Registers sind verschiedene Kenntnisse, Erfahrungen und Fertigkeiten notwendig. Ein *Register-Team* besteht in der Regel aus mehreren Fachexperten: einem *Projektleiter* zur Koordination und laufenden Überwachung der verschiedenen Handlungsstränge, mehreren *klinischen Beratern*, einem *Statistiker* für wissenschaftliche Auswertungen, einem *Datenmanager* zur Implementierung der Erhebungsformulare und Qualitätssicherung, einem *Informatiker* zur Umsetzung der IT-Infrastruktur sowie einem *Datenschutzbeauftragten* zur Wahrung der datenschutzrechtlichen Anforderungen und zum Entwurf des Datenschutzkonzepts. Kann kein Register-Team gebildet werden, sind externe Berater hinzuzuziehen. Anhaltspunkte hierfür liefert die infrastrukturelle Anbindung aus Protollpunkt 3.3.

An dieser Stelle im Registerprotokoll sind die Verantwortlichen namentlich zu nennen und mit ihren Aufgaben und Berechtigungen zu beschreiben.

6.2 Zeitplanung

Die Zeitplanung bezieht sich nicht nur auf die einzelnen Phasen der Registerentwicklung, sondern legt auch die zeitlichen Anforderungen während des Betriebs beispielsweise für Auswertungen, Publikationen und interne Qualitätsberichte fest.

Häufig unterschätzt wird dabei Aufwand und Dauer einer Testphase, um festzustellen, wie lange Dokumentationsvorgänge benötigen, wie aufwändig sie sind, und ob alle Fragestellungen entsprechend den zeitlichen Vorgaben beantwortet werden können.

6.3 Informationstechnische Werkzeuge

Es sind die soft- und hardwaretechnischen Werkzeuge zu beschreiben, die im Register eingesetzt werden. Dies beinhaltet die Nennung und Beschreibung der einzelnen Produkte, Systeme, Server und Kommunikationsschnittstellen.

7 Änderungen gegenüber der letzten Protokollversion

Klinische Register sind in der Regel langfristig ausgerichtet und nicht statisch. Daher können sich Rahmenbedingungen, Ziele, Anforderungen usw. im zeitlichen Verlauf verändern. Beispielsweise können sich neue Fragestellungen ergeben, bestehende an wissenschaftlicher Bedeutung verlieren, andere Datenquellen werden erschlossen oder neue technische Lösungen etabliert. Ein klinisches Register sollte auf diese neuen Situationen reagieren können. Notwendige Anpassungen und Korrekturen sind an dieser Stelle zusammenfassend aufzuführen und gegebenenfalls zu begründen. Das Registerprotokoll sollte zu jedem Zeitpunkt den aktuell gültigen Stand des Registers wiedergeben und eine nachvollziehbare Dokumentation der Änderungen gegenüber der letzten Register- / Protokollversion erlauben.

4.1.2 Hinweise zur Anwendung

Um eine systematische Registerplanung zu ermöglichen, sollten die einzelnen Protokollpunkte des Registerprotokolls fortlaufend, während der Entwicklungsphasen inhaltlich ausgearbeitet werden. Dies ermöglicht es, Entscheidungen reproduzierbar abzuleiten und schriftlich festzuhalten und schafft damit eine Diskussionsgrundlage für alle Beteiligten.

Ein frühes und klares Verständnis sekundärer Datenquellen und ihrer Nutzung ist entscheidend bei der Registerplanung, um Zielsetzungen, Anforderungen und schließlich das System selber geeignet herzuleiten. Entsprechend werden im Registerprotokoll die vorhandenen Datenquellen bereits in Protokollpunkt 3.2 eingeführt und können auf die Registerziele in Protokollpunkt 2.2 zurückwirken, sodass diese bei Bedarf angepasst werden können. Dieser Aufbau des Registerprotokolls erlaubt es, jedes erfasste Merkmal (Protokollpunkt 5.1) eindeutig als pro- oder retroektiv zu kennzeichnen und dabei zu prüfen, ob eine Erschließung aus sekundären Datenquellen überhaupt möglich (Qualität, Eigentümer usw.) oder ob eine problemorientierte Erfassung notwendig ist (Aufwand, Nutzen).

Werden klinische Register ausschließlich nach einem problemorientierten Vorgehen geplant, dann wird eine Vielzahl an Informationen, die in Routine- und wissenschaftlichen Datenquellen vorhanden ist, nicht genutzt, oder ist erneut und damit redundant zu erheben. Eine Kombination

aus problem- und datenorientiertem Ansatz impliziert, dass neben konkreten Problemen auch konkrete Datensätze und Datenquellen existieren, die in das Register integriert werden sollen. Abhängig von den Datenquellen und der geplanten Datennutzung kann sich die Zielsetzung des Registers gegenüber einem reinen problemorientierten Vorgehen verändern.

Zweckmäßig bei der Anwendung des Registerprotokolls kann es sein, dass nicht jeder Protokollpunkt schriftlich festgehalten wird, sondern dass an geeigneten Stellen auf externe Dokumente oder Systeme verwiesen wird. Dies vereinfacht es beispielsweise Anpassungen an Plausibilitätsregeln oder am Datenschutzkonzept vorzunehmen, ohne dass direkt eine Überarbeitung des gesamten Registerprotokolls notwendig wird. Ein Versionsmanagement analog zu Protokollpunkt 7 ist allerdings auch bei externen Dokumenten sinnvoll.

Die praktische Anwendung des Vorgehensmodells wurde am Beispiel des Registerprotokolls des klinischen Myelomregisters des Universitätsklinikums Heidelberg gezeigt und ist im Kapitel 5.1 beschrieben.

4.2 REFERENZ-IT-ARCHITEKTUR

Informationsverarbeitende Systeme sind elementare Bestandteile für die Umsetzung von Registern und unterstützen bei Erfassung, Speicherung, Qualitätssicherung und Kommunikation. Mit steigenden Anforderungen und längeren Laufzeiten erhöht sich die Komplexität der Systeme und bereits bei der Planung sind langfristig ausgelegte und flexible technische Komponenten zu beachten. Das im vorherigen Kapitel entwickelte Vorgehensmodell unterstützt die systematische Ableitung von Systemanforderungen an ein klinisches Register und berücksichtigt frühzeitig die Integration der informationstechnischen Systeme in die existierenden Systeme von Versorgung und Forschung.

Da sich die IT-Infrastrukturen klinischer Register in Abhängigkeit der eigenen Anforderungen und individuellen Umsetzung unterscheiden, wurde eine Referenz-IT-Architektur entwickelt, die bei Gestaltung, Auswahl und Realisierung der verschiedenen Systemkomponenten und Anwendungssystemen unterstützt und mehrere Implementierungsvarianten bietet. Zur praktischen Nutzung der Referenz-IT-Architektur wurden Entscheidungen, deren Konsequenzen und entsprechende Alternativen an verschiedenen Stellen herausgearbeitet und dargestellt. Der gesamte Prozess ist am Ende dieses Kapitels kompakt visualisiert.

4.2.1 Anwendungssysteme

Die Referenz-IT-Architektur eines klinischen Registers besteht aus drei Säulen, die sich hinsichtlich ihrer Planungs- und Realisierungsanforderungen unterscheiden, diese sind: die Systeme des klinischen Registers selbst, die Dokumentationssysteme der Versorgung und die Dokumentationssysteme der Forschung. Während die Systeme des klinischen Registers immer notwendig sind, kann der Umfang der anderen beiden Säulen in Abhängigkeit der Komplexität des aufzubauenden Registers und der Integration variieren. Abbildung 1 stellt die drei Säulen und die zugehörigen Anwendungssysteme grafisch dar.

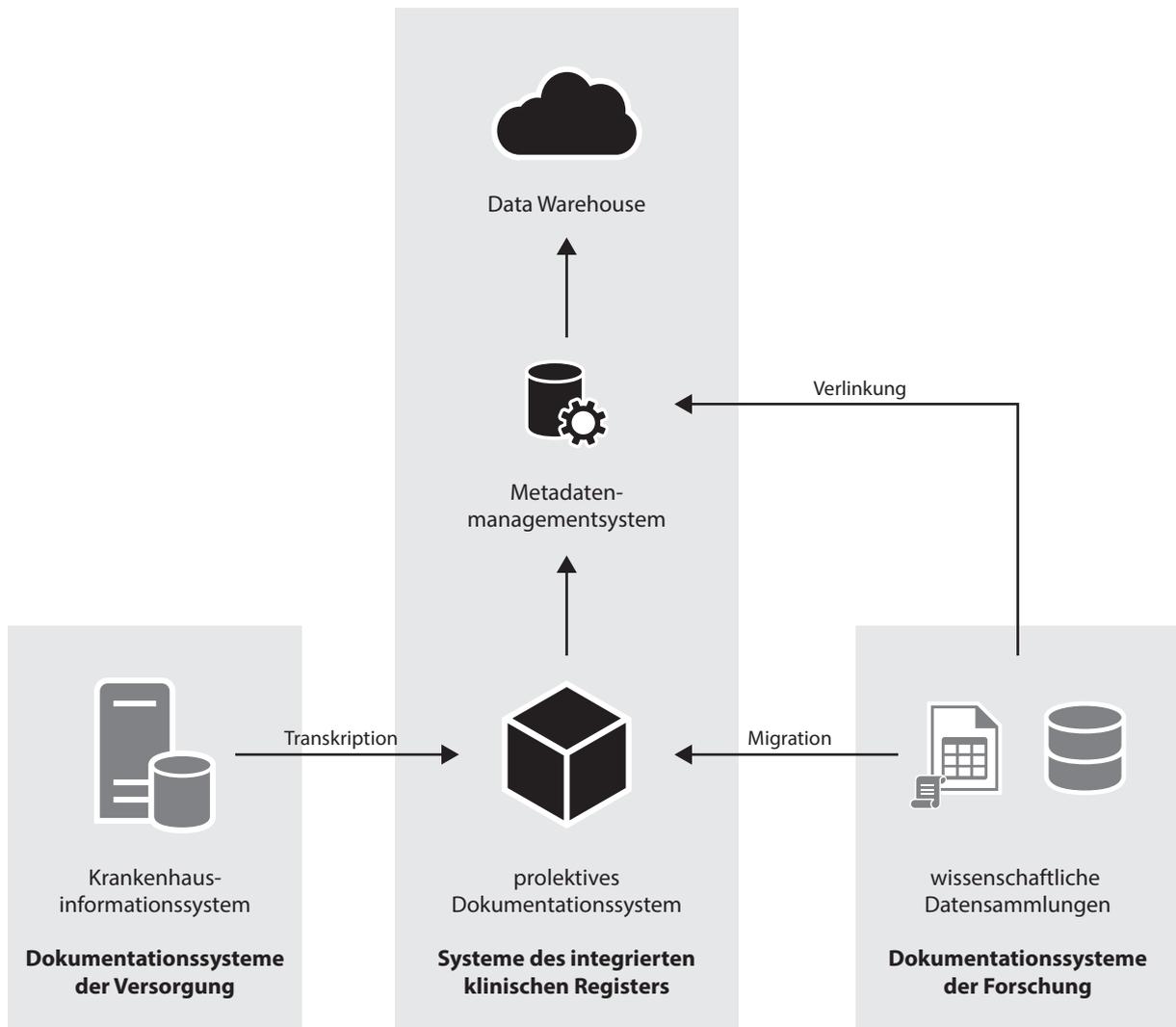


Abbildung 1: Referenz-IT-Architektur und zugehörige Anwendungssysteme.

Der Mittelpunkt der Referenz-IT-Architektur ist das elektronische Datenerfassungssystem zur Dokumentation der im Registerprotokoll spezifizierten Merkmale, die zur Beantwortung der wissenschaftlichen Fragestellungen notwendig sind. Ist keine Integration, und damit keine Übernahme retroaktiv vorhandener Daten vorgesehen, so besteht der informationstechnische Teil des klinischen Registers hauptsächlich aus diesem Datenerfassungssystem. Es kann daher auch als *prolektives Dokumentationssystem* verstanden werden, das im Falle eines neu aufzubauenden klinischen Registers mindestens realisiert und eingeführt werden muss. Im nachfolgenden Kapitel 4.2.2 wird beschrieben, welche Eigenschaften prolektive Dokumentationssysteme für klinische Register haben sollten, wie sich diese von einem Studiendatenmanagementsystem (SDMS) unterscheiden, nach welchen Kriterien die Systemauswahl erfolgen kann und wie eine praktikable datenschutzkonforme Lösung technisch realisierbar ist.

Klinische Register basieren in den meisten Fällen auf Informationen und Merkmalen, die in der Versorgungsroutine entstanden und in einem Patientenmanagementsystem (PMS) oder einem System zur Verwaltung der elektronischen Patientenakte (EPA-System) abgelegt sind. Zur Nutzung der klinischen Informationen ist eine Integration in die Dokumentationssysteme der Versorgung notwendig und eine Übernahme der klinischen Daten. Dieser Vorgang wird in dieser Arbeit als *Transkription* bezeichnet. Es entsteht eine (gewollte) Redundanz derselben Information, die sowohl den wissenschaftlichen Zielen des klinischen Registers als auch der Patientenversorgung dient. In Abhängigkeit der Systemeigenschaften und der zu integrierenden Merkmale kann die Transkription nach verschiedenen Verfahren erfolgen, die in Kapitel 4.2.3 näher erläutert werden.

In klinisch-wissenschaftlichen Versorgungseinrichtungen existieren häufig mehrere Dokumentationssysteme und stark strukturierte Datensammlungen, mit dem Ziel klinische Daten für wissenschaftliche Auswertungen zu erfassen. Bei der Planung eines klinischen Registers ist daher zu prüfen, ob sich wissenschaftliche Ausrichtungen überlappen oder ob Daten früherer Patienten vorliegen, die bei aktuellen Fragestellungen unterstützen. Nicht selten rührt die Motivation für den Aufbau eines klinischen Registers aus einer historischen wissenschaftlichen Datensammlung und die Integration der Daten in das neue prolektive Dokumentationssystem des Registers wird relevant. Abhängig davon, ob die Integration einmalig und dauerhaft oder wiederholend und temporär erfolgen soll, wird zwischen *Migration* und *Verlinkung* unterschieden. In beiden Fällen entsteht keine Redundanz, da die Daten entweder vollständig aus einer wissenschaftlichen Datensammlung in das prolektive Dokumentationssystem des klinischen Registers übernommen werden, oder die Zusammenführung nur für den Auswertungszweck erfolgt, die Daten ansonsten in ihren jeweiligen Dokumentationssystemen verbleiben. Nähere Informationen

zu der Integration wissenschaftlicher Dokumentationssysteme enthält Kapitel 4.2.4.

Das Ergebnis der Integration mit den Dokumentationssystemen der Versorgung und den verschiedenen Datensammlungen der Forschung kann in einem Data Warehouse dargestellt und angezeigt werden. Die Informationsintegration erfolgt mehrstufig und überwindet schrittweise technische, syntaktische und semantische Heterogenität zwischen den Systemen und stellt eine eindeutige Identifikation der Patienten sicher. Für welche Anwendungszwecke eines klinischen Registers ein Data Warehouse parallel zu dem prolektiven Dokumentationssystem zweckmäßig ist und wie die technische Realisierung der Informationsintegration als ETL-Prozess erfolgen kann, wird in Kapitel 4.2.5 beschrieben.

Unterstützend bei der Realisierung der unterschiedlichen Integrationsarten ist ein Metadatenmanagementsystem (MDMS). Analog zu einem Identitätsdatenmanagementsystem (IDMS), das Informationen über die Patienten eines klinischen Registers enthält, erlaubt ein MDMS die Abbildung semantischer Beziehungen zwischen Merkmalen unterschiedlicher Datensammlungen. Entsprechend lassen sich inhaltlich identische Merkmale beschreiben, Benennungs-, Wertebereichs- und Skalierungskonflikte darstellen und dadurch auflösen. Die semantische Interoperabilität zwischen den verschiedenen zu integrierenden Systemen wird dadurch sicher- bzw. hergestellt. Herleitung und Umsetzung eines MDMS gemäß des Standards *ISO / IEC 11179* zeigt Kapitel 4.2.6.

4.2.2 *Prolektives Dokumentationssystem*

Ein elektronisches Datenerfassungssystem, auch Electronic Data Capture (EDC) System genannt, ist ein rechnerunterstütztes System zur Erfassung und Speicherung klinischer Informationen in einem Datenbanksystem (DBS). Es besitzt eine grafische Benutzeroberfläche zur Dateneingabe und bildet die Benutzerschnittstelle zwischen den medizinischen Dokumentaren und der technischen Repräsentation der Daten. Aktuell werden im Gesundheitswesen EDC-Systeme insbesondere zur Dokumentation in klinischen Studien eingesetzt und sind daher meist auf deren spezifische Anforderungen hin optimiert. Ob sich ein EDC-System, das als SDMS konzipiert wurde, auch für ein klinisches Register eignet, hängt von der notwendigen Flexibilität bei der Abbildung des zu beobachtenden Wirklichkeitsausschnitts ab.

Allgemein sollte mit einem EDC-System die Umsetzung der im Registerprotokoll spezifizierten Datenhaltung, bestehend aus Registerschema (Protokollpunkt 5.1.1) und Merkmalskatalog (5.1.2), technisch möglich sein. Dies beinhaltet die Implementierung der identifizierten Domänen und zugeordneten Merkmale in Form von elektronischen Erfassungsf formularen (electronic Case Report Forms (eCRFs)). Zusätzlich sollten

Ereignisse (Events) definierbar sein, mit deren Hilfe Beziehungen zwischen Domänen und deren zeitliche Abfolge darstellbar sind. Durch die Zuordnung mehrerer eCRFs zu einem Event und eines eCRF zu mehreren Events kann eine wiederholende Datenerfassung zu unterschiedlichen Zeitpunkten realisiert werden. Auf diese Weise kann durch Kombination von Events und eCRFs das Behandlungs- und Therapieschema abgebildet werden. Abbildung 2 stellt diesen Zusammenhang beispielhaft dar.

Das Behandlungs- und Therapieschema ist bei klinischen Studien, insbesondere bei klinischen Interventionsstudien in der Regel fest vorgegeben und wird vorab von der Studiengruppe definiert. Dabei folgen alle

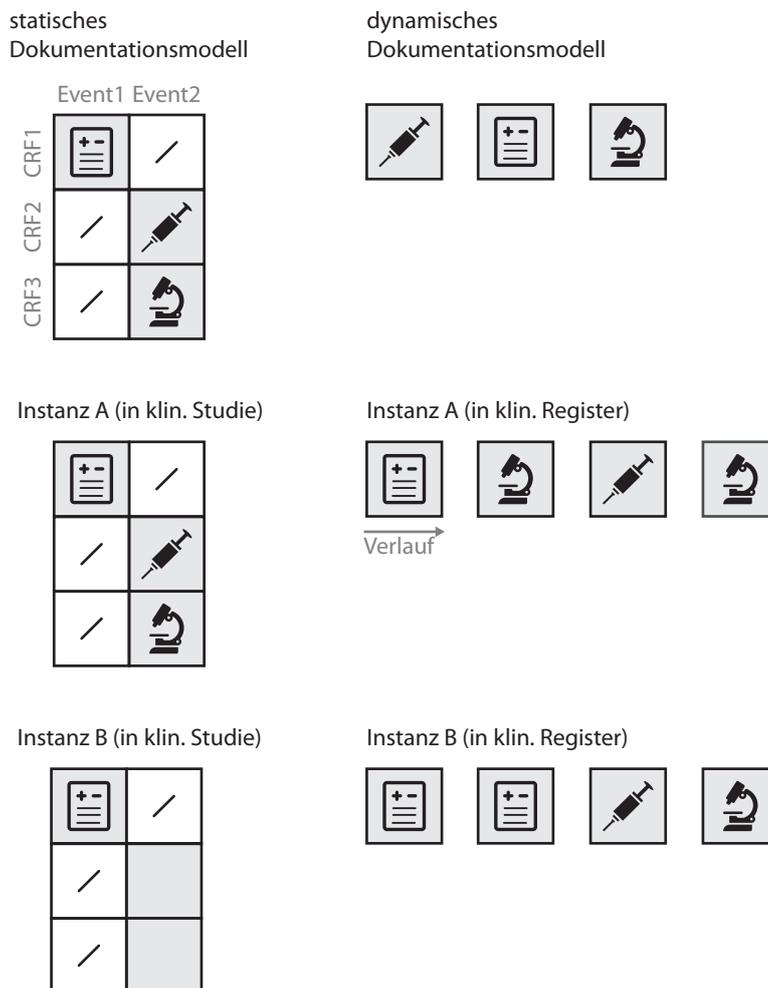


Abbildung 2: Gegenüberstellung verschiedener Dokumentationsmodelle zur Abbildung von Krankheits-, Behandlungs- und Therapieverläufen. Klinische Studien verwenden überwiegend das statische Dokumentationsmodell, klinische Register das dynamische.

in die klinische Studie eingeschlossenen Instanzen (Patienten) demselben Behandlungsschema. Die Zeitpunkte der Datenerhebung können aus diesem Schema verhältnismäßig einfach hergeleitet und als *statisches Dokumentationsmodell* im SDMS realisiert werden. Bei klinischen Registern und der nativen Beobachtung der Versorgungsrealität besteht allerdings die Zielsetzung nicht darin, ein Behandlungs- und Therapieschema vorzugeben oder zu standardisieren, sondern unterschiedlich behandelte und therapierte Patienten möglichst versorgungs- und realitätsnah abzubilden. Abhängig davon, wie individuell der zu beobachtende Wirklichkeitsausschnitt dargestellt werden soll und wie komplex dieser letztendlich ist, ist ein statisches Modell eventuell nicht geeignet, unterschiedliche Behandlungs- und Therapieverläufe adäquat abzubilden. Insbesondere bei Krebsbehandlungen sind oft verschiedene Therapieoptionen und -alternativen möglich, die den individuellen Gegebenheiten des Patienten angepasst, variabel kombiniert und wiederholt auftreten können. Sind diese Informationen relevant, bzw. soll die Versorgungsrealität wirklichkeitstreu dokumentiert werden, dann ist ein *dynamisches Dokumentationsmodell* einem statischen vorzuziehen. Dabei werden klinisch relevante Informationen zwar auch mit derselben Systematik definiert und erfasst wie in klinischen Studien üblich, aber die zeitliche Eventabfolge bleibt variabel. Da das Krankheits- und Therapiemodell vorab nicht definiert werden muss (oder aber nicht definiert werden kann), ist der Entwurf der eCRFs und die Dokumentation der klinischen Informationen einfacher, kann aber einen höheren (Nachbearbeitungs-)Aufwand bei späteren Auswertungen zur Folge haben. Von Auswertungsseite ist ein statisches Dokumentationsmodell immer vorzuziehen, von Dokumentationsseite in klinischen Registern ein dynamisches. Eine Übersicht der wesentlichen Eigenschaften und Unterschiede beider Modelle zeigt Tabelle 7.

Ob ein dynamisches oder ein statisches Dokumentationsmodell im konkreten Anwendungsfall zu realisieren ist, lässt sich während der Planungsphase des klinischen Registers über folgende Fragen feststellen:

- Haben alle Patienten denselben Behandlungs- und Therapieverlauf?
- Lassen sich alle Events vorab definieren?
- Sind alle Events einmalig (oder können sie wiederholt auftreten)?
- Gibt es eine feste Anzahl wiederholbarer Events (oder ist diese variabel)?
- Gibt es eine feste Reihenfolge der Events (oder ist diese variabel)?
- Lassen sich eCRFs den Events fest zuordnen?

Tabelle 7: Unterschiede zwischen einem statischen und einem dynamischen Dokumentationsmodell.

	STATISCH	DYNAMISCH
Ansatz	festes Behandlungs- und Therapie-schema; nicht schemakonforme Dokumentation problematisch	flexibles Behandlungs- und Therapieschema
Instanziierung	nicht vorgesehen; jede Instanz mit identischem Schema bzw. echter Teilmenge davon	möglich; jede Instanz kann individuelles Schema haben
Realisierung	bei <i>einfachem</i> Wirklichkeits-ausschnitt einfach realisierbar; bei <i>komplexem</i> Wirklichkeitsausschnitt Herleitung sehr aufwendig	relativ einfach herleitbar
Dokumentation der Versorgungsrealität	aufwändig, falls klinische Informationen und Events nicht auf Modell abbildbar (Einzelfall); Entscheidung erfolgt während Erfassung	einfacher, da klinische Informationen und Events auf Modell abbildbar sind
Auswertungen	einfacher, da klinische Informationen und Events bereits <i>aggregiert</i> vorliegen	aufwändig, da Instanzen, klinische Informationen und Events zunächst <i>aggregiert</i> werden müssen; Entscheidung erfolgt während Auswertung / Export
Datenqualität	in der Regel höher und einfacher zu gewährleisten (z. B. Vollständigkeit der Dokumentation)	in der Regel niedriger und aufwändig zu gewährleisten (z. B. Vollständigkeit der Dokumentation bleibt unklar)
Plausibilitätsprüfungen	einfach zwischen klinischen Informationen und Events realisierbar	komplex realisierbar, bei verschachtelten und eventübergreifenden Regeln

Können all diese Fragen mit *ja* beantwortet werden, dann besteht eine große Chance, dass ein statisches Dokumentationsmodell ausreichend ist, um den Wirklichkeitsausschnitt und damit den Behandlungs- und Therapieverlauf zu dokumentieren. Letztendlich hängt es von den im Registerprotokoll spezifizierten Systemanforderungen ab, welches Modell zu bevorzugen ist. Unabhängig davon, muss das wesentliche Kriterium bei Auswahl oder Implementierung eines Dokumentationssystems sein, dass sich das zu erfassende Krankheits-, Behandlungs- und Therapie-schema damit abbilden lässt. Erlaubt ein EDC-System beispielsweise nur statische Schemata, dann lassen sich damit keine individuellen Verläufe beschreiben und umgekehrt.

Planungsaspekte

Kommerzielle oder frei verfügbare EDC-Systeme sind in der Regel auf die Anforderungen klinischer Studien ausgelegt. Registerdatenmanagementsysteme (RDMS), d. h. spezifisch für klinische Register optimierte EDC-Systeme, sind nach aktueller Kenntnislage des Autors nicht weit verbreitet. Viele klinische Register nutzen daher Eigenentwicklungen auf Basis eines DBS mit einer (Web-)Oberfläche (Min et al. 2011; Plotnicki et al. 2013; Singh et al. 2012). Hierbei ist allerdings der zusätzliche zeitliche, finanzielle und personelle Aufwand zu berücksichtigen, der während der Aufbau- und Betriebsphase entsteht. Unabhängig davon, ob eine Eigenentwicklung oder die Auswahl und Parametrisierung eines EDC-Systems angestrebt wird, sind gewisse Eigenschaften notwendig, die ein RDMS haben sollte und die es von einem reinen SDMS unterscheiden. Nachfolgende Auflistung gibt einen Überblick typischer Eigenschaften eines RDMS. Produktspezifisch oder einzelfallabhängig sind allerdings Abweichungen möglich, sodass immer eine individuelle Prüfung notwendig wird, ob die jeweiligen Punkte den eigenen Systemanforderungen entsprechen.

MERKMALS-, CRF- UND EVENTDEFINITION Die zentrale Funktion aller EDC-Systeme ist die Definition von Merkmalen (Items) und deren Anordnung auf einem eCRF. Für jedes Merkmal sollten mindestens ein interner *Variablenname*, eine natürlichsprachige *Bezeichnung* für die Benutzeroberfläche und ein *Datentyp* definiert werden können. Einfache Datentypen für *Text*-, *Zahl*- und *Datumsangaben* sollten ebenso wie Wertemengen als *Einfach*- oder *Mehrfachauswahllisten* mit eindeutiger Codierung und Bezeichnung angelegt werden können. Zudem sollte auch die Definition von Events zur Beschreibung zeitlicher Verläufe und die Zuordnung ein oder mehrerer eCRFs zu einem Event möglich sein. Grafische Editoren, die die Umsetzung und Definition unterstützen, sind für Datenmanager intuitiver nutzbar als Skript- oder Programmiersprachen. Bei Eigenentwicklungen werden eCRFs gewöhnlich direkt in der Entwicklungsumgebung implementiert, wofür IT-Kenntnisse und Kenntnisse im Datenmanagement notwendig sind. Bei parametrisierbaren EDC-Systemen lassen sich eCRFs über *menschenlesbare* Sprachen oder grafische Editoren realisieren und erfordern weniger spezifische IT-Kenntnisse.

PLAUSIBILITÄTSREGELN Die Definition von Plausibilitäten auf Basis definierbarer Regeln ist zur Sicherstellung der Datenqualität notwendig. Grundsätzlich kann zwischen *harten* und *weichen* Regeln unterschieden werden. Harte Regeln sind meistens merkmalsbezogen, d. h. sie gelten für ein spezifisches Merkmal und stellen

beispielsweise die Richtigkeit des Datentyps sicher. Benutzereingaben, die gegen harte Regeln verstoßen, beispielsweise die Eingabe von Buchstaben in numerische Feldern, können nicht in dem DBS abgespeichert werden. Weiche Regeln hingegen informieren den Benutzer über Fehleingaben, erlauben aber deren Speicherung, wie beispielsweise das Unter- oder Überschreiten numerischer Wertebereichsgrenzen. Ob eine Regel als weich oder hart definiert wird, hängt von ihrer Wichtigkeit und Bedeutung ab. Zu viele harte Regeln können die Benutzereingabe beeinträchtigen und sollten daher möglichst gezielt eingesetzt werden. Neben *merkmalsbezogenen Regeln* sollten weiche Regeln auch *eCRF-bezogen* oder *eCRF-übergreifend* definiert werden können. Bei eCRF-bezogenen Regeln werden Plausibilitäten zwischen Merkmalen auf demselben eCRF geprüft, bei eCRF-übergreifenden zwischen Merkmalen verschiedener eCRFs. Plausibilitätsprüfungen können entweder in Echtzeit, d. h. bei Dateneingabe oder Speicherung, oder später, beispielsweise während eines klinischen Monitoring ausgeführt werden. Auch eine Kombination ist sinnvoll: so erlaubt eine Echtzeitprüfung die unmittelbare Korrektur während der Eingabe und retrospektive Prüfungen bestätigen die logische Konsistenz oder zeigen weitere Implausibilitäten auf. Ähnlich zu der Definition von eCRFs ermöglichen grafische Regeleditoren in EDC-Systemen häufig ein schnelles und einfaches Erstellen von Plausibilitätsprüfungen in Form einfacher logischer Ausdrücke. Bei Eigenentwicklungen werden diese oft in der Entwicklungssprache an verschiedenen Stellen der Benutzeroberfläche implementiert, was Nachvollziehbarkeit oder Anpassungen erschwert. SDMS verfügen meist zusätzlich über grafische Übersichten aller bestehender Konflikte, d. h. fehlgeschlagener Plausibilitätsprüfungen, denen dadurch gezielt nachgegangen werden kann.

In klinischen Studien ist es üblich, dass nur ein kleiner Teil der Plausibilitätsregeln in Echtzeit während der Dateneingabe erfolgt, und stattdessen die (logische) Konsistenz des exportierten Datensatzes mit Hilfe spezieller Statistik- und Auswertungssoftware sichergestellt wird. Dies gilt insbesondere bei eCRF-bezogenen und eCRF-übergreifenden Plausibilitäten. Entsprechend werden auch nicht alle Regeln im SDMS hinterlegt. Datenerhebung und Auswertung lassen sich aber bei klinischen Registern selten so strikt voneinander trennen wie in klinischen Studien und finden häufig parallel zueinander statt. Werden Daten exportiert, mit spezieller Software geprüft, im Fehlerfall korrigiert und wieder in das DBS zurückgeschrieben, dann kann dies aufwändig und fehleranfällig sein und den Dokumentationsfluss beeinträchtigen. Plausibilitätsprüfungen sollten daher in klinischen Registern direkt

im EDC-System definiert und auch in Echtzeit ausgeführt werden, anstatt eine spezielle Softwarelösung dafür zu nutzen. Daraus folgt, dass RDMS mit einer größeren Anzahl an Plausibilitätsregeln umgehen müssen als SDMS. Bei Systemparametrisierung oder Implementierung sind Datenverarbeitungsgeschwindigkeit und Systemleistung zu berücksichtigen, insbesondere, wenn viele merkmalsbezogene, eCRF-bezogene oder eCRF-übergreifende Plausibilitäten umgesetzt werden.

PATIENTENIDENTIFIKATION In einrichtungsbezogenen klinischen Registern erfolgt die Identifikation der Patienten im Regelfall über identifizierende Merkmale wie den *Namen*, das *Geburtsdatum* oder eine eindeutige, einrichtungsweite *Patienten-Identifikationsnummer*. In klinischen Studien werden meist *Pseudonyme* mit studienweiter Gültigkeit verwendet und externe Listen oder Systeme für die Speicherung, Verwaltung und Zuordnung zu den identifizierenden Daten (IDAT) genutzt, sogenannte IDMS. Viele SDMS beschränken sich daher allein auf den Umgang mit klinischen Daten (KDAT). Für den Dokumentationsfluss in klinischen Registern ist aber eine direkte Patientenidentifikation über identifizierende Merkmale hilfreich. Zwar besteht immer die Möglichkeit, ein eCRF mit identifizierenden Merkmalen zu definieren, allerdings erlauben nicht alle SDMS deren spezifische Abfrage, sodass eventuell die Suche eines Patienten über das Merkmal *Name* nicht realisiert werden kann. Verfügen EDC-Systeme über kein geeignetes Identitätsmanagement, müssen zusätzlich IDMS bei der Planung berücksichtigt werden.

Die Anforderungen eines klinischen Registers können über die eines IDMS hinausreichen. Insbesondere bei Langzeitregistern und der Organisation des Follow-Up können neben identifizierenden Merkmalen auch Adressdaten, Termine, niedergelassene Ärzte, Anschriften, Einverständniserklärungen, Widerrufe usw. relevant werden. Diese Informationen lassen sich nur selten mit Funktionen eines EDC-Systems abbilden¹ und spezifische Systeme oder Eigenentwicklungen werden notwendig. Um den Dokumentationsfluss dennoch einfach zu gestalten und datenschutzkonform IDAT von KDAT zu trennen, bietet sich beispielsweise eine Präsentationsintegration an.

¹ Relationale Beziehungen wie die M:N-Beziehung können mit EDC-Systemen nur umständlich realisiert werden, da alle Domänen derselben Informationseinheit zugeordnet werden. Verschiedene Informationseinheiten und deren Beziehung sind in der Regel nicht vorgesehen. Beispielsweise ist dies bei der Verwaltung von Patienten und ihrer behandelnden Ärzte relevant, wenn ein Patient mehrere Ärzte und ein Arzt mehrere Patienten haben kann.

AUDIT TRAIL Eine Audit Trail Funktion gehört zur guten klinischen Praxis (Good Clinical Practice; GCP) und wird von vielen SDMS unterstützt. Im engeren Sinn ist darunter die Protokollierung aller Benutzeraktionen wie die Eingabe von Daten oder deren Veränderung zu verstehen. Etwas weitergefasst lassen sich mit einem Audit Trail auch Systemvorgänge wie Datenexporte, Veränderungen an eCRFs oder Benutzerrechten registrieren und nachvollziehen. Häufig können neben einer reinen Berichtsfunktion Aktionen auch wiederholt ausgeführt oder rückgängig gemacht werden. So sollte beispielsweise nicht nur festgehalten werden, wann ein bestimmter Benutzer eine Merkmalsänderung durchgeführt hat, sondern im Idealfall auch welcher alte Wert wie überschrieben wurde. In klinischen Studien ist zudem die Angabe von Gründen bei Datenänderungen durch den Benutzer meist verpflichtend. Audit Trails können in klinischen Registern eine sinnvolle Ergänzung sein und eine gezielte Prüf-, Korrektur- und Wiederherstellungsmöglichkeit im Fehlerfall ermöglichen. Bei Eigenentwicklungen ist diese Funktionalität entsprechend zu berücksichtigen und einzuplanen. Kritisch könnte bei klinischen Registern mit Langzeitverläufen und der Beobachtung von Patienten über mehrere Therapiephasen hinweg der *Datenoverhead* sein, der bei einem Audit Trail entsteht. So werden zu jedem im DBS gespeicherten Merkmal automatisch weitere Informationen hinterlegt, die zusätzlichen Speicherplatz und Systemleistung erfordern. Wie umfangreich und wie lange die Informationen des Audit Trails zurückreichen sollen, ist im Einzelfall zu prüfen. Ein Kompromiss könnte beispielsweise die Protokollierung nur der jeweils letzten Änderung sein. Weiter zu berücksichtigen ist der Umgang mit Daten, die bei widerrufenem Patienteneinverständnis zu löschen sind, aber über ein Audit Trail prinzipiell wiederhergestellt werden können. Die erfordert häufig einen manuellen Eingriff in das DBS, um die entsprechenden Informationen zu entfernen.

VERSIONSKONTROLLE Die Versionskontrolle von eCRFs ist eine fortgeschrittene Eigenschaft von EDC-Systemen, die aber insbesondere bei klinischen Registern im Langzeitverlauf relevant werden kann. Klinische Register sind anders als klinische Studien selten auf einen primären Endpunkt hin optimiert, sondern erfassen die Versorgungsrealität mit all ihrer Dynamik. Folglich können sich im Verlauf Beobachtungsschwerpunkte verschieben, neue Forschungsfragen ergeben oder bestehende klinische Definitionen und Klassifikationen verändern. Entsprechend sollten EDC-Systeme für klinische Register eine höhere Flexibilität, Erweiterbarkeit und Adaptionfähigkeit aufweisen. So ist es in klinischen Registern

häufiger notwendig Merkmale oder eCRFs neu zu definieren oder bestehende anzupassen. Das Hinzufügen einzelner Merkmale oder ganzer eCRFs ist gewöhnlich wenig problematisch, da sie im DBS neu gespeichert und ergänzend auf der Benutzeroberfläche angezeigt werden können. Sind aber inhaltlich veraltete Merkmale zu entfernen oder in ihren Eigenschaften, z. B. in ihrem Wertebereich, zu verändern, dann ist sowohl der organisatorische als auch der technische Umgang zu klären, da auch veraltete Merkmale noch relevante Informationen enthalten können. Eine Versionskontrolle protokolliert nicht nur Änderungen an eCRFs und Merkmalen, sondern speichert Informationen sowohl instanz- als und versionspezifisch ab, wodurch verschiedene Instanzen (Patienten) unterschiedliche Versionen desselben Merkmals oder desselben eCRFs haben können². Steht keine Versionskontrolle zur Verfügung und lassen sich alte und neue Merkmale mit wenig Aufwand aufeinander abbilden, dann ist eine Datentransformation sinnvoll, die allerdings häufig nur auf Ebene des DBS möglich ist³. Als Alternative bleibt die Option, dasselbe Merkmal als *alt* und *neu* in EDC und DBS vorzusehen. Bei Eigenentwicklungen sind Anpassungen an Merkmalen, eCRFs und am DBS in der Regel nur durch den Entwickler möglich, eine Konsequenz, die beispielsweise bei Auftragsentwicklungen durch Unternehmen, Studenten oder Doktoranten zu berücksichtigen ist.

REPORT- UND BERICHTSFUNKTION Eine nützliche Funktion, die die meisten SDMS besitzen, ist es Abfragen (Queries) zu erzeugen und die Ergebnisse auf der Oberfläche darzustellen. Entsprechend sind Auswertungen nicht nur mit Hilfe spezieller Statistiksoftware und exportiertem Datensatz möglich, sondern auch im laufenden Betrieb, durch einen Datenmanager oder einen Dokumentar. Klinische oder wissenschaftliche Anfragen, beispielsweise zum Patientenkollektiv, lassen sich dadurch gegebenenfalls direkt beantworten. Verfügen EDC-Systeme über grafische Editoren, dann können Abfragen benutzerfreundlich erstellt und ohne datenbankspezifische Skript- oder Programmierkenntnisse realisiert werden. Werden (deskriptive) Ergebnisse zusätzlich noch automatisch

- 2 Existieren von demselben Merkmal verschiedene Versionen, beispielsweise *Geschlecht_1.0* {1, männlich; 2, weiblich} und *Geschlecht_1.1* {1, männlich; 2, weiblich; 3, unbekannt}, dann kann pro Instanz ausgewählt werden, ob die Erfassung mit Version 1.0 oder mit 1.1 erfolgen soll. Eine Transformation ist nicht notwendig.
- 3 Eine Erweiterung von Merkmalsausprägungen, beispielsweise *Geschlecht* {1, männlich; 2, weiblich} um eine dritte Option {3, unbekannt} ist meist trivial. Eine Reduktion, z. B. aus {1, männlich; 2, weiblich; 3, unbekannt} soll die dritte Option wegfällen, erfordert eine Prüfung und gegebenenfalls Nacherfassung aller inkonsistenten Fälle und deren Transformation. Alternativ ist auch eine (redundante) Speicherung derselben Information als *Geschlecht_alt* und *Geschlecht_neu* möglich.

visuell aufbereitet, können (einfache) statistische Auswertungen und Darstellungen unmittelbar im EDC-System erfolgen. Report- und Berichtsfunktionen erlauben auch häufig das Abspeichern und erneute Ausführen von Queries auf dem aktuellen Datenbestand, wodurch sich z. B. regelmäßige Berichte einmal definieren und wiederholt erzeugen lassen. Ein Spezialfall der Berichtsfunktion sind Fehlerberichte zur Sicherung von Datenqualität und Validität, die häufig nach Instanzen sortiert alle fehlgeschlagen Plausibilitäten auflisten und damit deren gezielte Korrektur ermöglichen. Eigenentwicklungen müssen, falls entsprechende Funktionen gewünscht werden, dahingehend erweitert werden.

BENUTZER- UND ZUGRIFFSMANAGEMENT Ebenfalls zum Standard vieler SDMS und auch in klinischen Registern relevant, ist ein Nutzer-, Rollen- und Rechtekonzept. Unerlässlich dabei ist die Nutzerauthentifizierung über Benutzername und Passwort. Rollen- und Rechteverwaltung sollten feingranular, am sinnvollsten auf Instanz-, eCRF- und Funktionsebene, definierbar sein. Beispielsweise sollten innerhalb eines multizentrischen Registers medizinische Dokumentare nur Zugriff auf Patienten des eigenen Zentrums haben, Statistiker keine identifizierenden Merkmale einsehen oder Praktikanten Datensätze nicht exportieren dürfen. Werden Rollen- oder Zugriffsrechte fest vom EDC-System vorgegeben, sind diese mit den eigenen Anforderungen abzugleichen. Differenzierte Sichten, d. h. an Benutzerrollen angepasste Benutzeroberflächen, sind hingegen meist optional.

DATENEXPORT UND -IMPORT RDMS sind selten geschlossene Systeme, sondern erfordern eine Möglichkeit klinische Informationen für Austausch, Qualitätssicherung und Auswertung freizugeben. Die einfachste Methode hierfür ist der Export. Mindestens sollte dieser in Standarddateiformaten wie CSV oder XML möglich sein, besser noch zusätzlich in spezifischen Formaten gängiger Statistiksoftware oder in international anerkannten (offenen) Standards wie CDISC (Clinical Data Interchange Standards Consortium). Lässt sich im EDC-System spezifizieren, welche Merkmale exportiert werden sollen, sodass neben Komplett- auch Teilexporte möglich sind, können weitere Zwischenverarbeitungsschritte bei der Freigabe der Daten an Kliniker oder Wissenschaftler vermieden und Merkmale unmittelbar und spezifisch für eine bestimmte Fragestellung bereitgestellt werden. Sinnvoll ist hier eine Kombination mit der Report- und Berichtsfunktion und dem gezielten Ausschluss irrelevanter, geschützter oder identifizierender Merkmale. Neben der Exportfunktionalität ist in klinischen Registern, und damit im Unterschied zu klinischen Studien, häufig auch die Übernahme von

Daten relevant, um beispielsweise extern erhobene oder qualitätsgeprüfte Merkmale effektiv zu verarbeiten. Zwar besteht immer die Möglichkeit Merkmale händisch in das EDC-System einzupflegen, schneller und sicherer ist aber deren Import insbesondere, wenn Daten bereits elektronisch und in einem EDC-kompatiblen Dateiformat vorliegen. Ergänzend zum Austausch klinisch relevanter Informationen auf Dateiebene, lässt sich über Systemschnittstellen eine direkte Anbindung und Kommunikation zwischen verschiedenen Dokumentationssystemen umsetzen. Schnittstellen oder Standards wie HL7 oder DICOM im Gesundheitswesen bieten zudem den Vorteil, dass sie branchenspezifisch weit verbreitet und einheitlich umgesetzt sind. Sehen EDC-Systeme die Anbindung externer Systeme vor, dann sind auch häufig geeignete Systemschnittstellen implementiert und nutzbar. SDMS sind darin allerdings häufig limitiert und eine Dateneingabe ist aus rechtlichen Gründen nur zugriffsberechtigten Benutzern vorbehalten. Klinische Register sind nicht durch Arzneimittel- (AMG) oder Medizinproduktegesetz (MPG) eingeschränkt und können verstärkt Schnittstellen zur Integration externer Daten und Systeme nutzen. Bei Auswahl und Prüfung eines EDC-Systems ist daher auf implementierte Systemschnittstellen und auf Kompatibilität zu vorhandenen Schnittstellen der Versorgungs- und Forschungsinformationssystemen zu achten. Verfügen EDC-Systeme über keine Schnittstellen und haben auch keine Exportmöglichkeit, dann sollte die Mindestforderung ein Direktzugriff auf das DBS sein. Dieser ist nicht nur maßgeblich für temporäre Export- und Integrationsvorgänge, sondern auch zur Daten- und Ausfallsicherung. Hinzu kommt, dass bei langen Registerlaufzeiten auch Technologie- und Paradigmenwechsel berücksichtigt werden müssen. Sind EDC und DBS *geschlossen*, lassen sich die darin dokumentierten Merkmale gegebenenfalls nur von Experten rekonstruieren oder sind erneut zu erfassen.

Insgesamt lassen sich die Eigenschaften eines RDMS gegenüber einem SDMS unter dem Aspekt *erhöhte Anpassbarkeit* zusammenfassen. Diese wird aufgrund längerer Laufzeiten, individueller Systemanforderungen und eines höheren Integrationsbedarfs in die bestehenden Informationssysteme von Versorgung und Forschung notwendig. So werden beispielsweise klinische Interventionsstudien meist einmalig initiiert und, sobald das Studienprotokoll und die eCRFs (nach mehreren Revisionen) freigegeben wurden, nicht mehr geändert. Entsprechend müssen auch weniger Änderungen und Anpassungen von dem eingesetzten SDMS unterstützt werden. Die Datenerfassung in klinischen Registern ist flexibler, wie beispielsweise die Unterscheidung von statischem und dynamischem Dokumentationsmodell zeigt, und erfordert, nicht nur während

der Planungs- und Aufbauphase, sondern auch im späteren Betrieb, ein EDC-System, das einfacher an neue inhaltliche, technische und organisatorische Anforderungen adaptiert werden kann, ohne einen Produkt- oder Technologiewechsel erforderlich zu machen.

Die Gegenüberstellung von Produkteigenschaften mit eigenen, aus dem Registerprotokoll systematisch abgeleiteten Systemanforderungen macht eine individuelle Prüfung und Gewichtung notwendig. Dabei sind Eigenschaften und der Funktionsumfang mit den eigenen Anforderungen abzugleichen, und falls nicht oder nur gering relevant, mit niedriger Priorität bei der Produktbewertung zu beurteilen. Zielführend bei einer individuellen Produktauswahl bzw. der Eigenentwicklung eines EDC-Systems ist ein strukturiertes und einheitliches Vorgehen nach zuvor festgelegten Kriterien. Mögliche Bewertungskriterien sind in Tabelle 8 dargestellt und orientieren sich an den oben aufgeführten Eigenschaften eines RDMS.

Tabelle 8: Kriterien zur Bewertung und Beurteilung der Eignung von Softwareprodukten zur elektronischen Datenerfassung für ein klinisches Register.

DATENMANAGEMENT
<input type="checkbox"/> Merkmalsdefinition <ul style="list-style-type: none"> <input type="checkbox"/> definierbarer Variablenname <input type="checkbox"/> natürlichsprachige Bezeichnung auf Oberfläche <input type="checkbox"/> verschiedene Datentypen <input type="checkbox"/> definierbare Wertemengen (mit Codierung, Bezeichnung) <input type="checkbox"/> Einfachauswahloptionen <input type="checkbox"/> Mehrfachauswahloptionen
<input type="checkbox"/> eCRF-Definition <ul style="list-style-type: none"> <input type="checkbox"/> Zuordnung mehrerer Merkmale zu einem eCRF <input type="checkbox"/> variable Wiederholung von Merkmalen auf eCRF <input type="checkbox"/> logische Unterteilungsmöglichkeiten (in Abschnitte / Gruppen)
<input type="checkbox"/> Eventdefinition <ul style="list-style-type: none"> <input type="checkbox"/> Zuordnung mehrere eCRFs zu einem Event <input type="checkbox"/> variable Wiederholung von Events <input type="checkbox"/> flexible Verschachtelung von Events
<input type="checkbox"/> grafischer Editor für Merkmals-, eCRF- und Eventdefinition
VALIDITÄTSPRÜFUNGEN
<input type="checkbox"/> definierbare Validitätsprüfungen

Tabelle 8: (Fortsetzung)

<input type="checkbox"/> harte Plausibilitätsregeln
<input type="checkbox"/> weiche Plausibilitätsregeln
<input type="checkbox"/> merkmalsbezogene Plausibilitätsregeln
<input type="checkbox"/> eCRF-bezogene Plausibilitätsregeln
<input type="checkbox"/> eCRF-übergreifende Plausibilitätsregeln
<input type="checkbox"/> definierbares Verhalten im Fehlerfall
<input type="checkbox"/> grafischer Editor zur Regeldefinition
<input type="checkbox"/> Echtzeit Plausibilitätsprüfungen während Dateneingabe
<input type="checkbox"/> Retrospektive Plausibilitätsprüfungen
<input type="checkbox"/> Auflistung fehlerhafter Plausibilitätsprüfungen (pro Instanz)
PATIENTENIDENTIFIKATION
<input type="checkbox"/> definierbare identifizierende Merkmale
<input type="checkbox"/> Instanzsuche über identifizierende Merkmale
<input type="checkbox"/> weitere definierbare identifizierende Merkmale (Adressdaten, Termine, Einverständnisse, etc.)
AUDIT TRAIL
<input type="checkbox"/> Protokollierung von Benutzeraktionen
<input type="checkbox"/> benutzerspezifische Begründung bei Änderungen und Löschungen
<input type="checkbox"/> Aktionen umkehrbar / wiederherstellbar
<input type="checkbox"/> Protokollierung von Systemvorgängen
<input type="checkbox"/> definierbare Anzahl protokollierter Aktionen
VERSIONSKONTROLLE
<input type="checkbox"/> verschiedene Versionen desselben Merkmals / eCRFs
<input type="checkbox"/> wählbare Merkmals- / eCRF-Version
<input type="checkbox"/> ungenutzte Merkmale / eCRFs bleiben im DBS
REPORT- UND BERICHTSFUNKTION
<input type="checkbox"/> definierbare Abfragen
<input type="checkbox"/> wiederholt ausführbare Abfragen
<input type="checkbox"/> grafischer Editor zur Abfragedefinition
<input type="checkbox"/> numerische Anzeige der Ergebnisse
<input type="checkbox"/> grafische Darstellung der Ergebnisse
BENUTZER- UND ZUGRIFFSMANAGEMENT
<input type="checkbox"/> Benutzerauthentifizierung

Tabelle 8: (Fortsetzung)

<input type="checkbox"/> definierbare Benutzerrechte
<input type="checkbox"/> Zugriffsrechte auf Instanzebene
<input type="checkbox"/> Zugriffsrechte auf eCRF-Ebene
<input type="checkbox"/> Zugriffsrechte auf Funktionsebene
<input type="checkbox"/> definierbare Benutzerrollen
<input type="checkbox"/> rollenspezifische Sichten auf Benutzeroberfläche
DATENEXPORT UND -IMPORT
<input type="checkbox"/> definierbarer Datenexport
<input type="checkbox"/> Standarddateiformate
<input type="checkbox"/> spezifische Dateiformate
<input type="checkbox"/> Datenimport über Standarddateiformate
<input type="checkbox"/> definierbarer Datenexport über Schnittstellen
<input type="checkbox"/> Datenimport über Schnittstellen
<input type="checkbox"/> direkter Zugriff auf DBS
<input type="checkbox"/> kompletter Datenexport
<input type="checkbox"/> kompletter Datenimport
SYSTEMARCHITEKTUR
<input type="checkbox"/> Client-Server-Infrastruktur
<input type="checkbox"/> (Web-)Oberfläche
<input type="checkbox"/> Datenbanksystem
<input type="checkbox"/> Plattformunabhängigkeit
<input type="checkbox"/> Verschlüsselung
PRODUKT- UND ENTWICKLERINFORMATIONEN
<input type="checkbox"/> Open Source
<input type="checkbox"/> technische Dokumentation
<input type="checkbox"/> Benutzerhandbuch
<input type="checkbox"/> Schulungsunterlagen
<input type="checkbox"/> regelmäßige Fehler- und Produktaktualisierung

Eine alternative Bewertungsmöglichkeit stellt die CIPROS Prüfliste von Lindörfer und Mansmann dar (Lindörfer u. Mansmann 2015).

Maßgeblich bei der Produktbewertung ist neben den Bewertungskriterien auch, dass bei mehreren EDC-Systemen stets dieselben Maßstäbe und Kriterien angelegt und nicht unterschiedliche Systeme verschieden bewertet werden. Dementsprechend sind die Bewertungskriterien gemäß den eigenen Systemanforderungen vorab zu gewichten und zu priorisie-

ren, beispielsweise prozentual oder dichotom mit *relevant* / *nicht relevant*. Vergleichbarkeit und eine zweckmäßige Produktauswahl oder Eigenentwicklung können nur darüber sichergestellt werden.

Realisierungsaspekte

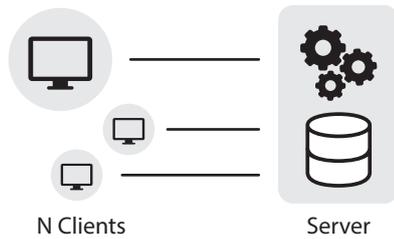
Bereits während der Produktbewertung oder Eigenentwicklung des EDC-Systems ist die Integration des prolektiven Dokumentationssystems in die bestehende IT-Infrastruktur der klinisch-wissenschaftlichen Versorgungseinrichtung notwendig. Ausgehend vom Systementwurf (Protokollpunkt 5) im Registerprotokoll und unter Berücksichtigung individueller technischer und organisatorischer Rahmenbedingungen ist die Systemarchitektur mit Kommunikationsdiensten, Hard- und Softwareausstattung festzulegen. Da die Eigenschaften des EDC-Systems direkten Einfluss auf die spezifizierte Systemarchitektur haben können, sind diese bei der Bewertung des EDC-Systems mit zu berücksichtigen (vergleiche Bewertungskriterien zur Systemarchitektur in Tabelle 8).

Entsprechend den in Kapitel 3.6 abgeleiteten und beschriebenen Anforderungen an ein integriertes klinisches Register ist die Systemarchitektur als Client-Server-Modell umzusetzen. Dabei werden Dienste und Anwendungen auf einem Server installiert und können von den verbundenen Clientrechnern genutzt werden. Bei einem klinischen Register ist das EDC-System zur Erfassung der klinischen Merkmale die bedeutendste auf dem Server installierte Anwendung, die von mehreren, über das Netzwerk angeschlossenen Clientrechnern parallel ausgeführt werden kann. Welche Systemarchitektur in einem konkreten Fall zu wählen und umzusetzen ist, hängt immer von den Eigenschaften des zu verwendenden bzw. zu implementierenden EDC-Systems ab. Eine schematische Darstellung der unterschiedlichen Systemarchitekturen zeigt Abbildung 3.

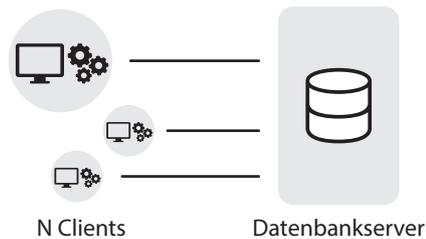
Wurde das EDC-System als *Webanwendung* konzipiert, dann finden Datenverarbeitung und Datenspeicherung auf dem Server statt, und Benutzereingaben und Ergebnispräsentation erfolgen über einen Webbrowser auf den Clientrechnern. Bei reinen *Datenbankanwendungen* hingegen sind Anwendungslogik und Oberfläche meist auf den Clients installiert und nur das DBS befindet sich auf dem Server. Ist eine Trennung zwischen Anwendungsoberfläche, Anwendungslogik und DBS möglich (und vom EDC-System vorgesehen), dann kann ein weiterer sogenannter *Anwendungsserver* zwischen Client und *Datenbankserver* geschaltet werden. Die Systemarchitektur und die Aufteilung zwischen Clients und Server(n) bestimmen die jeweiligen Leistungsanforderungen und den Ressourcenbedarf der einzelnen Systeme. Bestimmte Realisierungsaspekte sind aber dennoch zu berücksichtigen und nachfolgend beschrieben:

SERVERSYSTEM Ein Server ist im einfachsten Fall ein Computer auf dem eine Serversoftware installiert ist und der sich physisch nicht

Webanwendung (2-Schichten-Architektur)



Datebankanwendung (2-Schichten-Architektur)



Webanwendung (3-Schichten-Architektur)

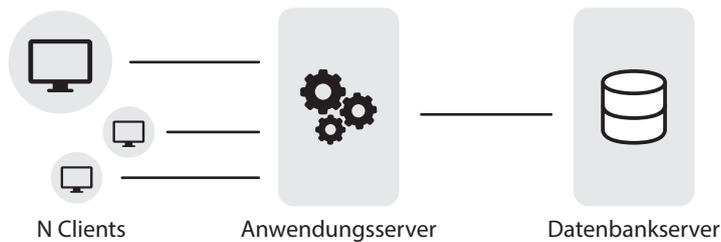


Abbildung 3: Schematische Gegenüberstellung verschiedener Client-Server-Modelle zur Realisierung des prolektiven Erfassungssystems des klinischen Registers.

von einem Clientrechner unterscheidet. Über die Serversoftware werden Dienste und Anwendungen wie EDC oder DBS den verbundenen Clientrechnern verfügbar gemacht. Entsprechende Systeme stehen meist in normalen Büroräumlichkeiten und sind hinsichtlich ihrer Ressourcen nicht auf Dauerbetrieb oder Ausfallsicherheit ausgelegt. Einen Arbeitsplatzrechner als *Server* zu nutzen sollte nur bei klinischen Registern mit geringer Verfügbarkeit, niedriger Nutzeranzahl und eingeschränkter Finanzierung in Erwägung gezogen werden.

Eine fortgeschrittene Variante sind Server, die bezüglich ihrer Hardwareausstattung, beispielsweise Festplatten, Prozessoranzahl, redundante Netzteile usw. auf Dauerbetrieb und hohe Zuverlässigkeit

ausgelegt sind. Solche physischen Server werden *Hosts* genannt und verfügen ebenfalls über eine Serversoftware, die Dienste und Anwendungen den Clients bereitstellt. Aus Sicht eines Clientrechners lässt sich nicht unterscheiden, ob eine Anwendung auf einem Host (physischen Server) oder einem normalen Computer (mit Serversoftware) ausgeführt wird. Hosts sind häufig zusammen mit anderen Hosts in separaten Gebäuden, Rechenzentren oder Serverparks mit eigener Klimatisierung und Energieversorgung untergebracht, wodurch sich Zugriffs- und Ausfallsicherung weiter erhöhen. Viele klinisch-wissenschaftliche Versorgungseinrichtungen verfügen über Rechenzentren, die als Serverstandort in Betracht gezogen werden sollten. Oftmals lassen sich dabei auch weitere Dienstleistungen und Ressourcen wie regelmäßige Datensicherungen nutzen. Soll ein Server in einem Rechenzentrum betrieben werden, sind zuvor die Nutzungsbedingungen zu prüfen und mit den eigenen Anforderungen abzugleichen: so können beispielsweise externe Zugriffe über das Internet, d. h. außerhalb des lokalen Netzwerks, untersagt sein oder identifizierende Daten dürfen nicht gespeichert werden. Entsprechende Einschränkungen sind bei der Systemarchitektur zu berücksichtigen.

Gewöhnlich verfügen Hostsysteme über genügend Rechen- und Leistungskapazität um mehrere Aufgaben parallel auszuführen. Eine empfohlene Möglichkeit diese sinnvoll einzusetzen ist die sogenannte *Servervirtualisierung*. Dabei wird ein physischer Server in mehrere virtuelle Server (VS) bzw. virtuelle Maschinen (VMs) aufgeteilt, die eigene, voneinander unabhängige Aufgaben und Anwendungen ausführen können. Zwar nutzen die verschiedenen VS dieselben Ressourcen und sind dadurch nicht vor einem Hardwaredefekt gesichert, erlauben aber eine unabhängige Administration, Wartung und Bedienung der verschiedenen Systeme. Beispielsweise lässt sich ein VS als Produktivumgebung des EDC-Systems nutzen und ein anderer als Entwicklungsumgebung, auf dem Weiterentwicklungen, Systemaktualisierungen oder eCRF-Entwürfe unabhängig vom Routinebetrieb getestet und validiert werden können, bevor sie in den Produktiveinsatz übergehen. VS sind insbesondere bei EDC und DBS, die verglichen zu anderen klinischen und bildgebenden Anwendungen *ressourcenschonend* sind, eine gute und kostengünstige Alternative zu mehreren dedizierten physischen Systemen. Ein weiterer Vorteil ist, dass sich viele Anwendungsszenarien und Architekturkonzepte zunächst als VS realisieren lassen, bevor physische Server eingesetzt werden.

Durch VS ist es nicht möglich Zuverlässigkeit, Verfügbarkeit und Sicherheit zu steigern, sodass hierfür zusätzliche physische Server

notwendig werden, die im Notfall den Routinebetrieb übernehmen und eine unmittelbare Weiterarbeit ermöglichen. Sind bei einem klinischen Register Dokumentation und Meldung zeitkritisch und Systemausfall-, Wartungs- und Weiterentwicklungszeiten hinnehmbar, können redundante Serversysteme eher mit niedriger Priorität beim Systementwurf eingestuft werden.

Für klinische Register ergeben sich je nach Anforderung verschiedene Konstellationen und Kombinationen virtueller und physischer Server. Sinnvoll, auch wenn datenschutzrechtlich gegebenenfalls nicht notwendig, erscheint immer ein Systementwurf, der eine Trennung identifizierender und klinischer Merkmale mindestens auf unterschiedlichen VS vorsieht. Gründe hierfür sind, dass für IDAT und KDAT optimierte EDC-Systeme ausgewählt bzw. implementiert und spezifische Zugriffsberechtigungen auf voneinander unabhängige DBS festgelegt werden können. Zusätzlich ist im Bedarfsfall eine physische Aufteilung auf mehrere Server einfacher möglich, als wenn alle Merkmale in einem DBS gespeichert sind. Sollen IDAT und KDAT zwar in getrennten EDC-Systemen verwaltet, aber dennoch dem Benutzer als einheitliche Oberfläche dargestellt werden, bietet sich ein weiterer (virtueller) Server zur Präsentationsintegration der verschiedenen EDC-Systeme an. Die doppelte Anzahl an Servern, also sechs, wird notwendig, wenn jeweils Produktiv- und Entwicklungsumgebung voneinander getrennt werden. Datensicherung sollte durch die angebotenen Dienstleistungen der Rechenzentren und mehrerer gespiegelter Festplatten sichergestellt werden. Weitere Server sind hierfür im Regelfall nicht erforderlich. Regelmäßige Sicherungen der VS und Backups der DBS auf externen Speichermedien sind eigeninitiativ und zusätzlich durchzuführen. Die Hochverfügbarkeit der Systeme, falls diese relevant ist, kann über einen weiteren physischen Server oder einen Rechnerverbund realisiert werden. Die einzelnen Realisierungsmöglichkeiten des Serversystems sind in Tabelle 9 zusammengefasst.

CLIENTRECHNER Clientrechner, oder auch Arbeitsplatzrechner genannt, sind herkömmliche Computer mit einem oder mehreren Bildschirmen, die zur Dateneingabe und Ergebnisausgabe genutzt werden. Die Rechner sollten über das Netzwerk der klinisch-wissenschaftlichen Versorgungseinrichtung mit dem Serversystem verbunden sein und die bereitgestellten Anwendungen ausführen können. Werden Clientrechner für weitere klinische Tätigkeiten wie Befundschreibung, Terminvereinbarung, Tabellenkalkulation oder Bildverarbeitung genutzt, sind gewöhnlich keine weiteren spezifischen Leistungsanforderungen zur Ausführung des prolekti-

ven Dokumentationssystem des klinischen Registers notwendig. Die erforderliche Softwareausstattung ergibt sich dabei aus dem genutzten und auf dem Server installierten EDC-System: beispielsweise wird bei einer Webanwendung ein Webbrowser und bei einer Datenbankanwendung das entsprechende lokal installierte Softwareprodukt benötigt.

Ist in einem klinischen Register die Speicherung und Verwaltung identifizierender Merkmale vorgesehen, dann sind spezifische EDC-Systeme, wie IDMS oder PMS, häufig besser dafür geeignet als reine SDMS. Die verschiedenen Systeme lassen sich jeweils den Anforderungen entsprechend gezielt auswählen und nutzen. Diese Systemtrennung führt allerdings auch zu einer Trennung identifizierender und klinischer Merkmale. Zusätzlich kann die Trennung eine geforderte Maßnahme zur Erfüllung datenschutzrechtlicher Anforderungen sein. Werden klinische und identifizierende Daten in unterschiedlichen EDC-Systemen gespeichert, dann wird der Patientenbezug häufig über ein gemeinsam genutztes Merkmal, das *Pseudonym*, hergestellt. Entsprechend ist immer ein Zugriff auf beide Systeme erforderlich, um einen klinischen Datensatz einem Patienten zuzuordnen. Erhöhte Daten- und Zugriffssicherheit sind die Folge, da nutzerspezifische Zugriffsberechtigungen bzw. Einschränkungen sowohl auf Server- als auch auf EDC-Ebene definiert werden können. Besteht allerdings für medizinische Mitarbeiter oder Dokumente ein direkter Behandlungsbezug, so offenbaren sich die Nachteile mehrerer EDC-Systeme: sollen beispielsweise klinische Merkmale zu einem Patienten im KDAT-EDC-System dokumentiert werden, so ist zunächst im IDMS das richtige Pseudonym zu dem Patienten zu identi-

Tabelle 9: Realisierungsmöglichkeiten des Serversystems des prolektiven Erfassungssystems des klinischen Registers.

SERVERSYSTEM	
<input type="checkbox"/>	Physischer oder virtueller Server mit Produktivumgebung des EDC-Systems
<input type="checkbox"/>	EDC-System für IDAT
<input type="checkbox"/>	EDC-System für KDAT
<input type="checkbox"/>	Präsentationsintegration IDAT-KDAT
<input type="checkbox"/>	Physischer oder virtueller Server mit Entwicklungsumgebung des EDC-Systems
<input type="checkbox"/>	EDC-System für IDAT
<input type="checkbox"/>	EDC-System für KDAT
<input type="checkbox"/>	Präsentationsintegration IDAT-KDAT
<input type="checkbox"/>	Physischer Server zur Ausfallsicherung der Produktivumgebung des EDC-Systems

fizieren, ehe im KDAT-EDC-System der pseudonymisierte Datensatz geöffnet und die Daten eingegeben werden können. Es besteht die Gefahr einer ineffizienten Dateneingabe und -ausgabe. Dies gilt insbesondere dann, wenn EDC-Systeme verschiedener Hersteller mit heterogenen Benutzeroberflächen und Strukturprinzipien genutzt werden. Eine Möglichkeit dedizierte EDC-Systeme einzusetzen und dennoch den Dokumentationsfluss optimal zu gestalten, ist die Präsentationsintegration über einen sogenannten *Portalserver*. Dieser zusätzliche (virtuelle) Server dient als Benutzerschnittstelle und ist spezifisch für die visuelle Integration mehrerer Komponenten, sogenannter *Portlets*, auf einer Webseite vorgesehen. Einzelne Portlets können für spezifische Aufgaben entwickelt werden und mehrere lassen sich gemeinsam auf einer Webseite zusammenstellen und anzeigen. Für die Integration eines IDMS und eines EDC-Systems mit gespeicherten KDAT sind entsprechend zwei Portlets, die die jeweiligen Anwendungsoberflächen kapseln, notwendig. Alternativ zu der nativen Benutzeroberfläche des IDMS bietet sich auch eine vereinfachte Darstellung der identifizierender Daten, beispielsweise in Form einer elektronischen Patientenliste (EPL) an. Sind nicht alle Merkmale des IDMS zur Patientenidentifikation und zur Eingabe der klinischen Daten relevant, wie beispielsweise Anschriften oder extern behandelnde Ärzte, dann ist es zweckmäßig, nur die relevanten Merkmale über eine EPL in einem Portlet anzuzeigen. Über Schnittstellen oder den direkten Zugriff auf das DBS des IDMS lassen sich die einzelnen Instanzen automatisch in einem EPL-Portlet aufführen und aktualisieren. Einem EPL-Portlet sollte insbesondere dann der Vorzug gegenüber einer Oberflächenkapselung des IDMS gegeben werden, wenn die Aufnahme neuer und die Änderung bestehender Instanzen ein von der Erfassung der klinischen Merkmale unabhängiger Arbeitsschritt ist. Erfolgen beispielsweise Rekrutierung oder Einschluss neuer Patienten durch andere Mitarbeiter, an unterschiedlichen Orten oder sind zeitlich versetzt, dann kann hierfür ausschließlich das IDMS ohne visuelle Integration genutzt werden. Über das EPL-Portlet werden die eingegebenen und aktualisierten identifizierenden Merkmale aus dem DBS des IDMS abgefragt und auf der Oberfläche des Portalservers dargestellt. Der Informationsaustausch zwischen EPL- und KDAT-Portlet erfolgt nachrichtenbasiert und erfordert standardisierte (und dokumentierte) Schnittstellen des KDAT-EDC-Systems. Wird beispielsweise ein direkter Instanzaufruf über eine URL unterstützt, dann lässt sich diese URL innerhalb des EPL-Portlets generieren und an das KDAT-Portlet weiterleiten. Der entsprechende Patient wird im KDAT-Portlet geöffnet und die Dateneingabe klinischer Merkmale kann erfolgen. Bedienung und Nutzung des KDAT-EDC-Systems innerhalb des Portlets bleiben unverändert zur nativen Anwendungsoberfläche. Die visuelle Integration identifizierender und klinischer Daten erfolgt ausschließlich clientseitig auf dem Arbeitsplatz-

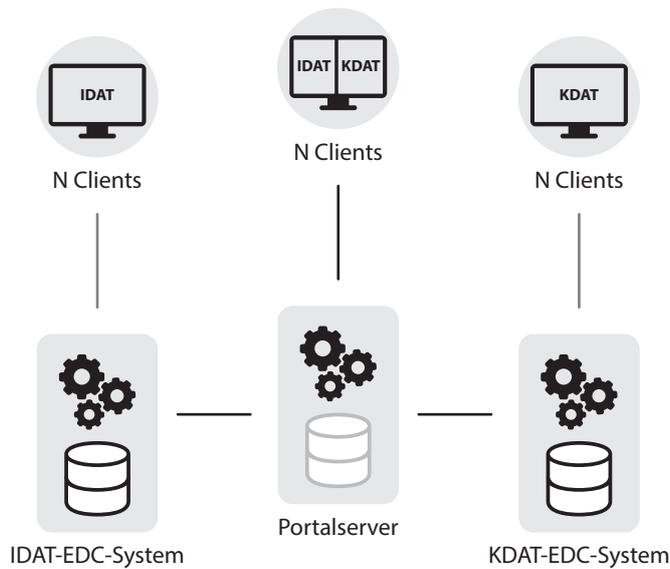


Abbildung 4: Präsentationsintegration zur visuellen Darstellung identifizierender und klinischer Merkmale dedizierter EDC-Systeme über einen Portalserver auf einer Weboberfläche.

rechner und nur innerhalb des Webbrowsers, EDC und DBS bleiben auf verschiedenen (virtuellen) Server getrennt. Zugriffsberechtigungen können daher entweder für das IDMS, das KDAT-EDC-System oder für den Portalserver mit identifizierenden und klinischen Merkmalen definiert werden. Die Funktionsweise dieses Ansatzes ist in Abbildung 4 schematisch dargestellt.

Mit der Bewertung eines prolektiven Dokumentationssystems geht auch häufig ein ausführlicher Produkttest einher. Die Testdurchführung sollte dabei alle relevanten Systemanforderungen abdecken und nach einem einheitlichen Testprotokoll erfolgen. Da Abbildung und Repräsentation des Wirklichkeitsausschnitts das wesentliche Kriterium für oder gegen ein EDC-System ist, sollten Merkmals-, eCRF- und Eventdefinition genau geprüft und Krankheits-, Behandlungs- und Therapieverläufe mit Real- und Testeingaben verifiziert werden. Dies gilt insbesondere bei dynamischen Dokumentationsmodellen, für die SDMS nur eingeschränkt geeignet sind. Ein entsprechendes Testprotokoll ist im Anhang A aufgeführt.

4.2.3 Integration mit Systemen der Versorgung

Klinische Register, die den Behandlungs- und Therapieverlauf von Patienten in der Routine beobachten, basieren in der Regel auch auf Dokumenten und Informationen, die in der Routine entstehen. Um eine

redundante Erhebung derselben Merkmale zu vermeiden, ist es sinnvoll auf die Informationen der Versorgungsroutine zurückzugreifen und diese für wissenschaftliche Fragestellungen in das RDMS zu integrieren. Voraussetzung hierfür ist, dass entsprechend dem Registerprotokoll, die retrolektiven Datenquellen der Versorgung bei Planung und Zieldefinition des klinischen Registers mitberücksichtigt und gemeinsam mit den neu zu erfassenden, prolektiven Merkmalen betrachtet werden.

Die Dokumentationssysteme der Versorgung setzen sich in der Regel aus vielen einzelnen Systemen zusammen, die alle Bestandteile des Krankenhausinformationssystems (KIS) sind. Der Schwerpunkt liegt dabei auf dem Patientendatenmanagement und auf der Dokumentation aller im Zusammenhang mit der Versorgung eines Patienten stehenden ärztlichen und pflegerischen Handlungen. Die Dokumentationspflicht für Ärzte ergibt sich dabei aus § 10 der (Muster-)Berufsordnung (MBO-Ä). Umgesetzt wird diese meistens über Patienten- oder Krankenakten, die zunehmend rechnerunterstützt verarbeitet und digital archiviert werden. Auch wenn die Infrastruktur eines KIS heterogen ist, fließen die meisten, die medizinische Versorgung eines individuellen Patienten betreffenden Informationen in die (elektronische) Patientenakte des EPA-Systems. Sie bildet den Krankheits-, Behandlungs- und Therapieverlauf unter Realbedingungen ab und ist inhaltlich in hohem Maße als Datenquelle für ein klinisches Register geeignet. Die strukturelle Eignung der klinischen Informationen ist allerdings nicht immer gegeben und muss merkmalspezifisch geprüft und gegebenenfalls hergestellt werden.

Sollen Routinedaten aus dem KIS in das Dokumentationssystem des klinischen Registers übernommen werden, dann ist die verwendete Integrationsart die *Transkription*. Dabei werden Daten aus den Systemen der Versorgung in das klinische Register übertragen. Da die Informationen weiterhin im Quellsystem vorhanden bleiben, liegen sie redundant vor. Hinzu kommt, dass nicht alle verfügbaren Informationen der Routineversorgung für die wissenschaftlichen Fragestellungen des klinischen Registers relevant sind, der Datenqualität genügen oder sich aufgrund ihrer Struktur eignen. Beispielsweise können Dokumente unstrukturiert, fortlaufend und in natürlicher Sprache verfasst sein, unvollständig im KIS vorliegen oder Befunde noch nicht bestätigt sein. Entsprechend kann auch nur ein kleiner Teil der digitalen Informationen aus einem KIS automatisch weiterverarbeitet werden.

Die Aufgaben bei der Transkription lassen sich so zusammenfassen, dass die notwendigen Informationen aus Dokumenten des KIS selektiert, aufgearbeitet und strukturiert in das EDC-System des klinischen Registers übertragen werden. Abhängig von den Eigenschaften der Daten sind verschiedene Formen der Transkription unterscheidbar: liegen Informationen nur papierbasiert oder digitalisiert vor, dann ist eine Transkription auch nur manuell durch medizinische Dokumentare möglich. Bei digital

verfügbaren Informationen ist nach dem Grad der Strukturierung zu differenzieren und zu prüfen, ob Merkmale direkt abgeleitet werden können oder zunächst aus Textpassage extrahiert werden müssen.

Aspekte bei Planung und Realisierung der Transkription als Integrationsart zwischen den Dokumentationssystemen der Versorgung und dem Dokumentationssystem des klinischen Registers werden in den folgenden Abschnitten beschrieben.

Planungsaspekte

Die systematische Planung eines klinischen Registers sieht im Registerprotokoll die Spezifikation der vorhandenen Routinedatenquellen und die Herleitung von Merkmalen zur Beantwortung wissenschaftlicher Fragestellungen vor. Das Ergebnis ist eine eindeutige Zuordnung von sekundärer Datenquelle und relevantem Merkmal, beispielsweise, dass aus Arztbriefen das *Diagnosestadium* dokumentiert werden kann, oder dass aus Laborbefunden der *Calciumwert* erfasst wird. Neben dieser semantischen und logischen Relation sind im weiteren Verlauf der Planung die Eigenschaften der Routinedatenquellen und der Routinedaten zu bestimmen und wie informationstechnisch auf die darin enthaltenen Informationen zugegriffen werden kann.

EIGENSCHAFTEN DER ROUTINEDATENQUELLEN Ein KIS ist selten auf rechnerunterstützte Komponenten beschränkt und umfasst häufig auch konventionelle Systeme und Archivierungsmethoden. Entsprechend sind nicht alle in einer Patientenakte enthaltenen Dokumente und Datenquellen elektronisch verfügbar, sondern können auch papierbasiert oder digitalisiert bzw. eingescannt vorliegen. Heutzutage werden Informationen innerhalb einer klinisch-wissenschaftlichen Versorgungseinrichtung häufig digital erzeugt und verarbeitet. Externe Kommunikation, beispielsweise mit niedergelassenen Fachärzten oder mit staatlichen Behörden erfolgt allerdings meist papierbasiert. Einrichtungsabhängig werden externe Dokumente entweder konventionell abgelegt und archiviert oder digitalisiert und einer elektronischen Patientenakte angefügt. Sind Dokumente elektronisch nicht verfügbar, enthalten aber relevante Informationen, so sind administrative Abläufe und die Datenerfassung so zu organisieren, dass eine Verarbeitung durch medizinische Dokumentare des klinischen Registers ermöglicht wird (siehe Protollpunkte 5.2.1). Ist die Dokumentation und/oder Meldung zeitkritisch, kann es sinnvoll sein, wenn externe Dokumente vor Digitalisierung und Speicherung zunächst an das klinische Register zur Erfassung weitergeleitet werden. Bei papierbasierten Dokumenten ist die manuelle Transkription durch medizinische

Dokumentare die einzige Möglichkeit, Informationen zu selektieren und in das EDC-System des klinischen Registers zu übertragen. Ähnliches gilt bei eingescannten Dokumenten, die rechtssicher und unveränderbar in digitalen Archivierungssystemen abgelegt sind. Ist eine optische Texterkennung (Optical Character Recognition; OCR) und Konvertierung in digitale Zeichenketten nicht zulässig oder wird vom Archivsystem nicht unterstützt, dann ist eine elektronische Weiterverarbeitung auch nicht möglich. Daten und Informationen, die innerhalb der eigenen Versorgungseinrichtung erhoben, erfasst und übermittelt werden, sind häufig die einzigen digital verfügbaren Datenquellen der Routine. Neben einer manuellen Transkription, die immer möglich ist, sind automatische Verfahren zu berücksichtigen, die sich hinsichtlich der Eigenschaften der Routinedaten und den Zugriffsmöglichkeiten auf das KIS unterscheiden.

EIGENSCHAFTEN DER ROUTINEDATEN Sind Routinedatenquellen digital im KIS verfügbar, dann ist zu prüfen, ob die Informationen in Freitexten enthalten sind oder bereits direkt als ableitbares Merkmal vorliegen. KIS unterstützen häufig beide Ansätze und sehen Dokumente mit starken und schwachen strukturellen Vorgaben vor. Neben einer Strukturierung auf Gliederungsebene lässt sich auch der Inhalt unterschiedlich stark vorgeben: eine *Diagnose* beispielsweise kann frei und mit verschiedenen Variationsmöglichkeiten beschrieben oder systematisch nach Haupt- und Nebendiagnose getrennt und mit standardisiertem Klassifikationssystem dokumentiert werden. Je stärker Gliederung und inhaltliche Festlegungen vorgegeben sind, desto einfacher die Weiterverarbeitung einzelner Informationen, sowohl manuell als auch elektronisch. Werden Daten strukturiert im KIS abgefragt und merkmalsbasiert gespeichert, dann liegen sie häufig auch in dieser Form im DBS des KIS vor und können gegebenenfalls automatisch weiterverarbeitet werden. Meist ist dies bei Aufnahmeformularen, Laborbefunden oder Abrechnungsinformationen der Fall. In klinischen Befunden und Arztbriefen werden hingegen häufig Freitexte verwendet, die zusätzliche Freiheiten bei der Erfassung komplexer Sachverhalte bieten und nur eingeschränkt elektronisch weiter verarbeitbar sind. Sind Informationen nur von medizinischen Dokumentaren bestimmbar oder müssen unter inhaltlicher Abwägung mehrerer Textpassagen interpretiert werden, dann bieten automatische Analyseverfahren selten eine ähnliche Trefferquote und Genauigkeit wie eine manuelle Erfassung. Liegen aber Informationen digital als Merkmal vor oder sind aus Freitextpassagen herleitbar, dann lassen sich manuelle Erfassungsarbeiten reduzieren und Merkmale auto-

matisch aus dem KIS in das Dokumentationssystem des Registers übernehmen.

ZUGRIFFSMÖGLICHKEITEN AUF DAS KIS Der Zugriff auf die Dokumentationssysteme der Versorgung erfolgt in der Regel über *Benutzerschnittstellen*, die zur Ein- und Ausgabe klinischer Daten genutzt werden, und immer, die erforderlichen Nutzerberechtigungen vorausgesetzt, eine manuelle Transkription von Informationen aus dem KIS in das EDC-System des klinischen Registers ermöglichen. Zusätzlich verfügen die einzelnen Systeme eines KIS häufig über *Softwareschnittstellen*, die den elektronischen Daten- und Nachrichtenaustausch untereinander realisieren. Da ein direkter Zugriff auf das DBS einer Versorgungseinrichtung selten möglich ist, bleiben Schnittstellen und *Datenexporte* häufig die einzigen Möglichkeiten des Zugriffs auf die digital verfügbaren Informationen der Versorgung. Die im Gesundheitswesen am häufigsten genutzten Schnittstellenstandards sind *HL7* und *DICOM*. Durch den Austausch von HL7-Nachrichten beispielsweise werden Informationen, die in einem System erfasst und eingegeben wurden, an ein anderes System zur Weiterverarbeitung übermittelt. Zusätzliche Systeme zum Empfang der Nachrichten lassen sich in diesen Prozess einbinden, erfordern allerdings eine Anpassung des Kommunikationsservers und setzen damit die Genehmigung und Befürwortung des informationstechnischen oder geschäftsführenden Leiters der Versorgungseinrichtung voraus. Sind Daten relevant, die bereits im KIS vorliegen, dann kann eine nachrichtenbasierte Übertragung nicht mehr erfolgen und der Zugriff auf digital gespeicherte Informationen ist nur über Exportfunktionen möglich.

In Abhängigkeit der Routinedatenquellen, der Routinedaten und deren Zugriffsmöglichkeiten lässt sich die Transkription unterschiedlich realisieren. Aspekte, die dabei zu berücksichtigen sind, sind nachfolgend aufgeführt.

Realisierungsaspekte

Die Integration von Informationen und Merkmalen aus der Versorgungsroutine in das prolektive Dokumentationssystem des klinischen Registers ist in den meisten Fällen nur über eine Transkription möglich, bei der Daten aus dem KIS in das klinische Register übertragen werden. Die Eigenschaften des KIS, der Routinedatenquellen und der darin enthaltenen Routinedaten bestimmen, ob als Transkriptionsart eine *manuelle*, ein *Text Mining* oder ein *ETL-Prozess* umgesetzt werden kann. Tabelle 10 fasst die jeweiligen Voraussetzungen zusammen.

Tabelle 10: Voraussetzungen zur Realisierung der drei Transkriptionsarten *manuell*, *Text Mining* und *ETL* zur Integration von Informationen aus den Dokumentationssystemen der Versorgung in das prolektive Dokumentationssystem des klinischen Registers.

ART DER TRANSKRIPTION	EIGENSCHAFTEN DER ROUTINE- DATENQUELLE	EIGENSCHAFTEN DER ROUTINE- DATEN	ZUGRIFFS- MÖGLICHKEITEN AUF DAS KIS
manuell	papierbasiert digitalisiert digital	unstrukturiert strukturiert	Benutzerschnittstelle
Text Mining	digitalisiert (mit OCR) digital	unstrukturiert	Benutzerschnittstelle Datenbanksystem
ETL	digital	strukturiert	Softwareschnittstellen Exportfunktion

Manuelle Transkription

Die manuelle Transkription ist die meist genutzte Transkriptionsart bei klinischen Registern, da bis auf Zugriffsberechtigungen keine weiteren informationstechnischen Voraussetzungen erfüllt sein müssen. Während bei papierbasierten Routinedatenquellen eine manuelle Übertragung die einzige Alternative darstellt, Informationen zu ermitteln und strukturiert in das EDC-System des klinischen Registers einzugeben, wird sie auch bei digitalen und digitalisierten Datenquellen relevant, wenn bis auf Benutzerschnittstellen keine weiteren Zugriffsmöglichkeiten auf das KIS angeboten werden. Die Qualität einer manuellen Transkription hängt maßgeblich mit der Ausbildung, Schulung und dem fachlichen Verständnis eines medizinischen Dokumentars oder Dokumentationsassistenten zusammen. Sind mehrere Dokumentare an einem manuellen Transkriptionsprozess beteiligt, dann sind verbindliche Vorgehensbeschreibungen (Standard Operation Procedures; SOPs), Benutzer- und Eingabehandbücher zweckmäßig, um eine einheitliche Datenerfassung zu ermöglichen. Schriftlich sollte mindestens für jedes zu transkribierende Merkmal festgehalten werden, in welcher Routinedatenquelle die betreffende Information enthalten ist und wie sie in das prolektive Dokumentationssystem übertragen werden soll. Tabelle 11 zeigt ein Beispiel zur eindeutigen Beschreibung und semantischen Zuordnung einer Information im KIS mit dem korrespondierenden Merkmal im klinischen Register.

Neben Vorgehensanweisungen für den Regelfall ist in klinischen Registern auch der Umgang mit Ausnahmen, Sonderfällen, fehlenden, unbekanntem und falschen Informationen vorzusehen. Da die ärztliche Dokumentation während der Versorgungsroutine häufig unter großem Zeitdruck entsteht, sind fehler- und lückenhafte Befunde und Arztbriefe nicht auszuschließen (Püschmann et al. 2006). Werden in einem klinischen

Tabelle 11: Eindeutige Beschreibung und semantische Zuordnung einer Information im KIS mit dem korrespondierenden Merkmal im klinischen Register bei der manuellen Transkription.

BESCHREIBUNG	BEDEUTUNG	BEISPIEL
Merkmal	Bezeichnung des Merkmals im Dokumentationssystem des klinischen Registers. Ggfs. der Variablenname.	Calciumwert
Klinische Information	Bezeichnung der klinischen Information im KIS, welche dem Merkmal im klinischen Register entspricht.	Calciumwert, Kalziumwert, CA
Routine-datenquelle	Bezeichnung der Datenquelle oder des Dokuments im KIS, die die klinische Information enthält.	Arztbrief
Abschnitt	Bezeichnung des Abschnitts in der Routinedatenquelle, falls diese in mehrere Gliederungspunkte unterteilt sind.	Laborwerte
Ggfs. Syntax	Syntaktische Beschreibung der klinischen Information	CA Calciumwert Einheit
Beispiel	Textausschnitt, Bild oder Screenshot der klinischen Information.	ALB 46.8 g/l CA 2.4 mmol/l HB 13.9 g/dl
Ggfs. Transformationsregel	Vorschrift zur Umwandlung der klinischen Information in das korrespondierende Merkmal.	mg/dl = mmol/l * 18.0182

Register Routinedatenquellen genutzt und Unzulänglichkeiten während der wissenschaftlichen Erfassung festgestellt, beispielsweise wenn klinische Informationen nicht übernommen werden können oder gegen Plausibilitätsregeln verstoßen, sind spezifische Abläufe zur Rückkopplung mit der Versorgungsroutine erforderlich. Die manuelle Transkription kann in so einem Fall als zusätzliche Maßnahme zur Qualitätssicherung und -verbesserung der ärztlichen Dokumentation verstanden werden. Selbst wenn Vorgehensbeschreibungen vorhanden sind, bleibt die Dokumentation ein zeit- und personalaufwändiger Prozess, und menschliche Interpretations- oder Aufmerksamkeitsfehler sind kaum vermeidbar.

Eine manuelle Transkription wird, wie die anderen Transkriptionsarten auch, immer durch ein bestimmtes Ereignis ausgelöst. Beispielsweise erst wenn ein Routinedokument im KIS vorhanden ist, kann der Dokumentationsvorgang beginnen. Problematisch wird diese Verfahrensweise dann, wenn Dokumentaren das auslösende Ereignis nicht mitgeteilt wird, oder diese wiederholend die Verfügbarkeit einer relevanten Information im KIS nachprüfen müssen. Dies führt zu langen Wiederholungsschleifen

und erschwert zeitnahe Erfassungen. Entsprechend sollten auch bei einer manuellen Transkription automatisierte Prozesse und Maßnahmen etabliert werden, die das Vorhandensein neuer Informationen an das klinische Register melden. Umsetzen lässt sich dies beispielsweise über Softwareschnittstellen, die keine Daten übermitteln, sondern über das Ereignis informieren, oder über eine direkte Kommunikation zwischen klinischem Urheber und medizinischem Dokumentar.

Text Mining

Eine Möglichkeit den Transkriptionsprozess automatisiert zu unterstützen ist das Text Mining, bei dem mit Hilfe computerlinguistischer Methoden des Natural Language Processings (NLP) versucht wird, natürlichsprachige Texte algorithmisch so zu verarbeiten und zu erfassen, dass Bedeutungsstrukturen erkannt werden. Ein Teilbereich des Text Mining ist das Information Retrieval (IR) und das gezielte Suchen, Extrahieren und Weiterverarbeiten von Informationen in un- oder schwachstrukturierten Textpassagen.

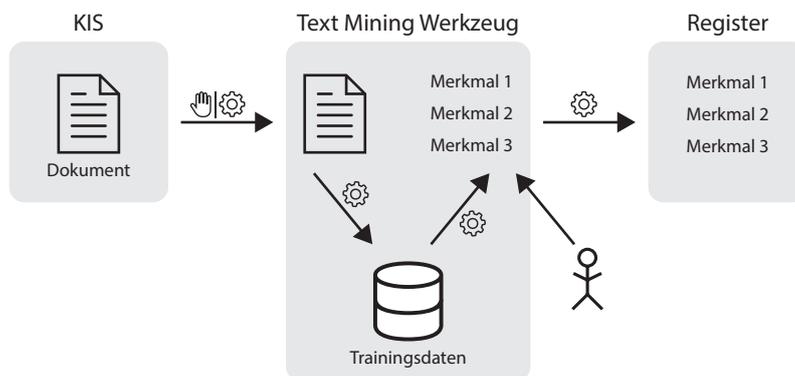


Abbildung 5: Automatisierte Transkription relevanter Informationen in un- oder schwachstrukturierten Textpassagen über das Text Mining zur Integration der Dokumentationssysteme der Versorgung.

Die eingesetzten Methoden beim IR sind überwiegend Klassifikatoren, die vordefinierte Informationen erkennen und in Kategorien einordnen. Voraussetzung für ein IR sind *Trainingsdatensätze*, die klinische Informationen und korrespondierende Merkmale enthalten und so in Beziehung zueinander setzen, dass Zusammenhänge algorithmisch feststellbar und automatisch erkannt werden können. Die Güte der Klassifikation hängt bei einem IR von der Qualität der klinischen Informationen und des Trainingsdatensatzes ab, die meist Abkürzungen, Rechtschreibfehler und uneinheitlich genutzte Strukturelemente aufweisen. Zur Aufbereitung klinischer Textpassagen vor der eigentlichen Klassifikation sind daher häufig Vorprozessierungen erforderlich. Obwohl IR ein etabliertes Verfahren ist,

müssen kommerzielle oder freie Produkte immer für den Einzelfall angepasst oder selbst entwickelt werden und erfordern individuelle Lösungen zur Anbindung an die Informationssysteme der Versorgung und des klinischen Registers. Ist der elektronische Zugriff auf das KIS nicht oder nur eingeschränkt möglich, dann kann beispielsweise ein halbautomatisches Verfahren umgesetzt werden, bei dem Textpassagen von medizinischen Dokumenten in ein *Text Mining Werkzeug* kopiert, der IR-Prozess ausgeführt, Fehler manuell korrigiert und die extrahierten Merkmale anschließend über Schnittstellen in das klinische Register übertragen werden. Auf diese Weise ließe sich eine schnelle und qualitätsgeprüfte Dokumentation umsetzen und die manuellen Korrekturen zur Verbesserung der Trainingsdaten und der Klassifikatoren nutzen.

Ein Text Mining zur Integration unstrukturierter Informationen aus Arztbriefen ist als Anwendungsbeispiel der Referenz-IT-Architektur im Kapitel 5.4 beschrieben.

ETL-Prozess

Die automatische Verarbeitung klinischer Informationen kann über sogenannte ETL-Prozesse (Extract, Transform, Load) umgesetzt werden. Dabei werden Daten aus einem System exportiert, mithilfe definierter Transformationsregeln umgewandelt und in ein anderes System geladen. Aufgrund der Einschränkung auf digital-strukturierte Informationen lässt sich mit ETL-Prozessen häufig nur ein Teilbereich der wissenschaftlichen Dokumentation abbilden und externe Kommunikation mit konventionellen oder digitalisierten Dokumenten nur über eine manuelle Transkription erfassen.

Die Umsetzung eines ETL-Prozesses erfordert den Zugriff auf die digital verfügbaren Informationen im KIS und wird in der Regel über *Softwareschnittstellen* oder *Exportfunktionen* realisiert. Der Unterschied besteht in der Extraktion der klinischen Informationen: während bei Softwareschnittstellen Daten in Form elektronischer Nachrichten weitergeleitet werden, ist der Informationsträger bei einem Export eine Datei. Bei einer Schnittstellenkommunikation wird der Nachrichtenaustausch

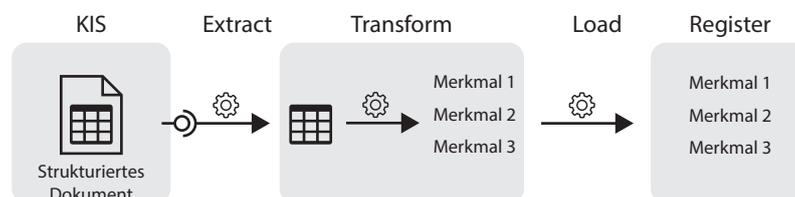


Abbildung 6: Automatisierte Transkription strukturierter Informationen über Softwareschnittstellen in das prolektive Dokumentationssystem des klinischen Registers.

zudem automatisch durch ein Ereignis initiiert und liefert nur aktuelle Informationen. Ist das EDC-Registersystem zum Zeitpunkt des auslösenden Ereignisses nicht verfügbar oder kann eine Nachricht nicht verarbeiten, kommt es zu einem Datenverlust. Zusätzlich werden pro Tag zahlreiche Nachrichten innerhalb des KIS generiert und versendet. Viele davon sind für die wissenschaftlichen Fragestellungen eines klinischen Register nicht relevant oder enthalten maschinell erzeugte Rohwerte, ohne Plausibilitätsprüfungen oder ärztliche Bestätigung. Dies birgt die Gefahr, dass zu viele Informationen oder solche mit fehlender fachlicher Verifizierung in das klinische Register integriert werden. Beispielsweise wenn eine Laboruntersuchung wiederholt durchgeführt werden muss, sind auch die nicht konklusiven Befunde über Schnittstellen automatisch in dem klinischen Register enthalten, aber weder klinisch noch wissenschaftlich nutzbar. Bei der Exportfunktion kann die Extraktion manuell angestoßen oder zeitlich festgelegt werden, sodass Häufigkeit und Informationsmenge besser steuerbar und kontrollierbar werden. Auch lassen sich bereits vorhandene Daten früherer Patienten exportieren und in das klinische Register integrieren, was bei einer Schnittstellenkommunikation häufig nicht mehr möglich ist, da Nachrichten nach erfolgreicher Übermittlung zwischen den Systemen des KIS gelöscht werden.

Liegen die extrahierten Daten als Nachricht oder Datei vor, dann kann die nachfolgende Prozessierung identisch ablaufen: Aufgabe des Transformationsschrittes ist es, die relevanten Informationen zu selektieren und auf das Schema und den Merkmalskatalog des klinischen Registers abzubilden, sodass ein Import möglich wird. Der Vorteil eines automatischen Verfahrens ist, dass Informationen nicht redundant erhoben und trotzdem (zeitgleich) in mehreren Systemen vorliegen können und, Übertragungsfehler ausgeschlossen, konsistent zum KIS sind. Im Einzelfall, nachrichten- und exportabhängig ist allerdings zu prüfen, ob bestimmte Informationen vollautomatisch in das klinische Register integriert werden sollen oder ob eine manuelle Erfassung nicht doch die sinnvollere Alternative darstellt. Eine Kombination aus automatischer Verarbeitung, wie der Variablenzuordnung oder Einheitenrechnung, und manueller

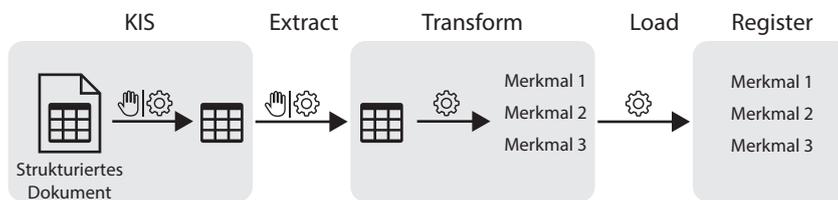


Abbildung 7: Automatisierte Transkription strukturierter Informationen über Exportfunktionen des KIS in das prolektive Dokumentationssystem des klinischen Registers.

Zwischenschritte zum Auslösen eines Exports oder zur Bestätigung des Imports können eine bewusste und zielgerichtete Transkription unterstützen.

Ein ETL-Prozess zur Integration strukturierter Laborbefunde wurde als Anwendungsbeispiel der Referenz-IT-Architektur entwickelt und ist im Kapitel 5.3 aufgeführt.

4.2.4 *Integration mit Systemen der Forschung*

Das Ziel einer Versorgungseinrichtung im klinisch-wissenschaftlichen Bereich besteht meist darin, klinische Daten der individuellen Patientenversorgung so zu erfassen und zu strukturieren, dass patientenübergreifende Aussagen ableitbar sind. Möglich wird dies durch verschiedene Forschungsmethoden wie klinische Studien oder systematische Erhebungen. Abhängig von einer konkreten wissenschaftlichen Fragestellung, den Rahmenbedingungen oder dem erwarteten Umfang, werden in der Regel mehrere Forschungsmethoden parallel angewendet und Daten aktiv in unterschiedlichen Dokumentations-, EDC- oder Datenbanksystemen erfasst. Zusätzlich ist anzunehmen, dass neben aktueller klinischer Forschung weitere Datensammlungen zu bereits abgeschlossenen und ausgewerteten Fragestellungen innerhalb einer klinisch-wissenschaftlichen Forschungseinrichtung existieren. Die in historischen und aktuellen Forschungsdatenquellen enthaltenen Informationen können auch für ein klinisches Register relevant sein und übergreifende Auswertungen, Kohorten-, Wirksamkeits- oder Überlebenszeitanalysen ermöglichen. Überschneiden sich Fragestellungen, Patientenkollektive oder Merkmale zwischen wissenschaftlicher Datensammlung und klinischem Register, dann ist eine Datenintegration einer redundanten Erhebung vorzuziehen. Dies gilt auch, wenn ein klinisches Register auf einer existierenden, aber nicht mehr zeitgemäßen wissenschaftlichen Dokumentation aufbauen soll. Bleiben Altdaten und historische Fallbeschreibungen unberücksichtigt, dann ist die wissenschaftliche Aussagekraft und der Nutzen eines klinischen Registers zu Beginn als gering einzuschätzen und gewinnt erst im fortgeschrittenen Betriebsstadium mit größer werdender Fallbasis an Bedeutung. Werden Altdaten stattdessen frühzeitig integriert, steigt der Informationsgehalt des klinischen Registers und gemeinsame Auswertungen historischer und neuerfasster Fälle werden möglich.

Die Herausforderung bei der Integration prolektiver Registerdokumentation mit retrolektiven Forschungsdatenquellen besteht darin, die Interoperabilität zwischen den verschiedenen Systemen sicherzustellen. Da wissenschaftliche Datensammlungen bereits existieren, ist es zweckmäßig während der systematischen Registerplanung die vorhandenen retrolektiven Daten hinsichtlich ihrer Eignung zu prüfen und prolektive Merkmalsdefinitionen entsprechend anzupassen. Erfolgt dies nicht, dann

ist von heterogenen Systemen auszugehen und Zusammenführungen oder Auswertungen sind nicht oder nur mit zusätzlichem Aufwand realisierbar.

In der klinischen Forschung werden verschiedene Dokumentationssysteme für die Erhebung, Erfassung und Speicherung wissenschaftlicher Daten genutzt. Das Dokumentationssystem bildet entweder selbst die Datensammlung ab oder verfügt über spezifische Eingabeoberflächen mit separatem DBS. Beispielsweise erfolgt bei einer Tabellenkalkulation die Eingabe unmittelbar in ein Spreadsheet, während bei einem EDC-System Erfassungsbögen auf einer Oberfläche einer Datenbank vorgeschaltet sind. Für den technischen Integrationsprozess mit den prolektiven klinischen Registerdaten ist allerdings weniger das wissenschaftliche Dokumentationssystem von Bedeutung, sondern vielmehr die dahinterstehende Datenquelle, die in der Regel stark strukturiert und merkmalsbasiert vorliegt. Aus organisatorischer Sicht sind vor der Integration die Nutzungsberechtigungen der Forschungsdatenquellen zu prüfen. Wurden Informationen zweckgebunden erhoben, wie in klinischen Studien üblich, dann ist eine Sekundärnutzung, die über den Forschungskontext der klinischen Studie hinausgeht, nicht zulässig. Zusätzliche Ethikvoten und/oder Einverständniserklärungen werden notwendig, die eine Zusammenführung und gemeinsame Auswertung mit den prolektiven Registerdaten vorsehen.

Analog zu der Integration mit den Dokumentationssystemen der Versorgung besteht auch bei wissenschaftlichen Datensammlungen immer die Möglichkeit einer Transkription. Dabei werden Merkmale von einem System in das prolektive Dokumentationssystem des Registers überführt und damit selbst Teil von diesem, bleiben aber weiterhin redundant in dem Quelldatensystem vorhanden. Während dies bei den Dokumentationssystemen der Versorgung zweckmäßig ist, da dieselben klinischen Informationen spezifisch für die Patientenversorgung als auch für wissenschaftliche Fragestellungen erhoben und erfasst werden können, führt dies bei Forschungsdatenquellen zu erhöhten Aufwänden, da die Daten mindestens stark strukturiert vorliegen. Entsprechend kann die Integration mit den Dokumentationssystemen der Forschung in höherem Maße automatisiert erfolgen. Als Integrationsart ist abhängig von den Eigenschaften der Forschungsdatenquelle, der Wiederholhäufigkeit und der Dauerhaftigkeit zwischen *Migration* und *Verlinkung* zu unterscheiden. Während das Ziel der Migration darin besteht eine Forschungsdatenquelle ganz oder teilweise abzulösen und die Merkmale ganz oder teilweise zu übernehmen (und anschließend aus dem Quelldatensystem zu entfernen), werden bei der Verlinkung die Merkmale für wissenschaftliche Auswertungen zusammengeführt, bleiben aber weiterhin in ihren separaten Dokumentationssystemen erhalten. Relevante Aspekte, die bei der

Planung und Umsetzung beider Integrationsarten zu berücksichtigen sind, werden in den folgenden beiden Abschnitten beschrieben.

Planungsaspekte

Entsprechend dem Vorgehensmodell zur systematischen Planung eines klinischen Registers wird frühzeitig geprüft, welche Forschungsdatenquellen relevante Informationen, Merkmale oder Instanzen zur Beantwortung prolektiver Fragestellungen enthalten. Wurden geeignete Datensammlungen identifiziert, ist anschließend ihre Verfügbarkeit und Eignung hinsichtlich datenschutzrechtlicher, technischer und semantischer Voraussetzungen zu prüfen. Aus Sicht eines klinischen Registers haben Forschungsdatenquellen immer einen retrolektiven Charakter und ihre Merkmale und Instanzen wurden zweckgebunden für eigene Fragestellungen und Auswertungen erhoben und erfasst, die heterogen zu denen eines klinischen Registers sein können. Unabdingbar für einen Integrationsprozess ist daher die rechtliche Erlaubnis zur Datennutzung, der informationstechnische Zugriff auf die Datenquelle und die semantische Interoperabilität zwischen den Systemen. Lässt sich die semantische Interoperabilität nicht während der Planungsphase eines klinischen Registers her- bzw. sicherstellen, Instanzen nicht eindeutig identifizieren oder sind retrolektive Definitionen für prolektive Fragestellungen nicht geeignet, beispielsweise da sie nicht aufeinander abbildbar sind oder Merkmale unterschiedliche Bedeutungsstrukturen haben, dann ist die Forschungsdatenquelle bzw. Teile davon für eine Integration mit den prolektiven Registerdaten ungeeignet. Der im Registerprotokoll spezifizierte Merkmalskatalog (Protokollpunkt 5.1.2) sollte daher so gestaltet sein, dass bei zu integrierenden Merkmalen eindeutig auf die enthaltene Datenquelle verwiesen und die Eigenschaften konsistent bzw. mit Transformations- und Abbildungsregeln syntaktisch ableitbar sind.

Das Ziel einer Integration mit den Dokumentationssystemen der Forschung ist es, den Informationsgehalt zu erhöhen und neue Erkenntnisse herzuleiten, die mit den einzelnen Systemen und Datensammlungen nicht möglich sind. Ein Informationsgewinn lässt sich dabei nur so realisieren, indem zu bestehenden Instanzen neue Merkmale oder zu bestehenden Merkmalen neue Instanzen hinzugefügt werden. Datensammlungen sind am einfachsten als Tabelle mit einer horizontalen und einer vertikalen Achse zu beschreiben: in horizontaler Richtung ist typischerweise eine Menge an Merkmalen, in vertikaler Richtung eine Menge an Instanzen abgebildet. Bei einer Integration gilt es zunächst festzustellen, ob Informationen zu denselben Instanzen oder zu denselben bzw. zu ähnlichen und aufeinander abbildbaren Merkmalen in den Datensammlungen enthalten sind. Liegen beispielsweise unterschiedliche Merkmale zu denselben Instanzen in verschiedenen Datensammlungen vor, dann erfolgt

bei einer Integration ein Informationsgewinn in horizontaler Richtung durch Zusammenführung (Vereinigung) der Merkmale. In der Praxis tritt dies insbesondere dann auf, wenn klinische Daten von Patienten mit administrativen, molekularbiologischen oder weiteren Gesundheitsdaten korreliert werden. Sind hingegen in zwei Datensammlungen unterschiedliche Kollektive mit denselben Merkmalen vorhanden, dann wird durch eine Integration die Menge an Instanzen in vertikaler Richtung erweitert. Abbildungen 8 und 9 stellen diesen Zusammenhang grafisch dar.

Ein Integrationsprozess erfolgt immer zweistufig: in einem ersten Schritt wird die Menge an übereinstimmenden, ähnlichen oder identischen Merkmalen oder Instanzen bestimmt, bevor in einem zweiten die jeweils andere Menge hinzugefügt wird. Voraussetzung ist die semantische Interoperabilität der Merkmale und die eindeutige Identifikation der Instanzen. Beide lassen sich während Planung und Aufbau eines klinischen Registers auch ohne informationstechnische Unterstützung sicherstellen, beispielsweise indem einheitliche Merkmalsdefinitionen und Identifikatoren genutzt werden, erfordern in der Praxis aber häufig spezifische Werkzeuge und Methoden zur Realisierung. Zweckmäßig ist daher

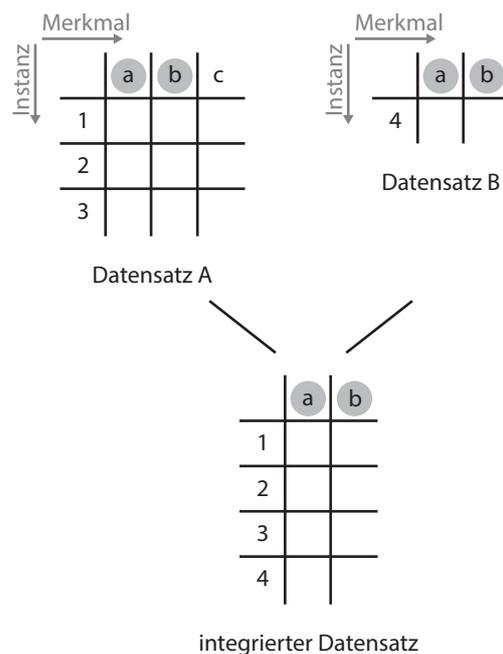


Abbildung 8: Integration in vertikaler Richtung: zu derselben Menge an Merkmalen (Schnittmenge) werden neue Instanzen integriert (Vereinigungsmenge). Beispielsweise bei der Integration zweier Patientenkollektive mit denselben (oder aufeinander abbildbaren) Merkmalen.

die frühzeitige Kenntnis, ob eine Integration mit Forschungsdatenquellen notwendig ist, und falls ja, ob dadurch zusätzliche Instanzen und/oder Merkmale integriert werden sollen, wovon wiederum abhängig ist, ob eine semantische Interoperabilität und/oder eindeutige Identifikation vorauszusetzen ist.

Die Integration eines prolektiven Dokumentationssystems mit einer retrolektiven wissenschaftlichen Datensammlung kann auf zwei Arten erfolgen: entweder über eine *Migration* oder eine *Verlinkung*. Der wesentliche Unterschied besteht darin, ob der Aufbau eines klinischen Registers die Nutzung der Forschungsdatenquelle beeinträchtigt und ob diese dann weiterhin relevant ist. Im Regelfall bleibt eine Forschungsdatenquelle auch bei bestehendem klinischen Register in einem eigenen und unabhängigen Dokumentationssystem und die Integration wird nur für den Zweck der wissenschaftlichen Auswertung durchgeführt. Die Integrationsart hierfür wird als Verlinkung bezeichnet und kann einmalig, aber auch wiederholend erfolgen und resultiert in einem temporären, zusammengeführten Datensatz, mit zusätzlichen Instanzen oder Merkmalen. Soll hingegen ein klinisches Register auf einer etablierten wissenschaftlichen Dokumentation aufbauen, dann bietet sich als Integrationsart die Migration der historischen Daten unmittelbar in das Dokumentationssystem des klinischen Registers an. Der wesentliche Vorteil der Migration ist, dass

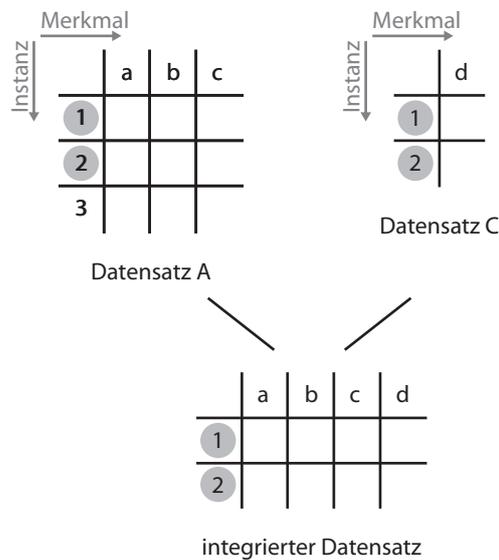


Abbildung 9: Integration in horizontaler Richtung: zu derselben Menge an Instanzen (Schnittmenge) in zwei Datensätzen werden neue Merkmale integriert (Vereinigungsmenge). Beispielsweise bei der Integration klinischer und administrativer Daten.

die historischen Fälle nach der Integration in demselben System weiterdokumentiert werden können, wie die prolektiven, und nicht mehrere Dokumentationssysteme parallel genutzt und gepflegt werden müssen. Zudem enthält das klinische Register bereits in der Anfangsphase des Betriebs die wissenschaftliche Fallbasis der klinisch-wissenschaftlichen Versorgungseinrichtung und kann diese unmittelbar für Abfragen und Auswertungen nutzen. Im Gegensatz zu einer Verlinkung ist eine Migration einmalig, löst das bestehende Dokumentationssystem ab und die integrierten Daten gehen dauerhaft in das klinische Register über. Da die zu entwerfenden Integrationsprozesse bei beiden Arten identisch sind, ist es von der weiteren Nutzung der Forschungsdatenquelle abhängig, ob diese dauerhaft in das klinische Register migriert oder wiederholend mit den prolektiven Registerdaten verlinkt werden soll. Ebenfalls für eine Entscheidung relevant ist die Qualität und Konsistenz der wissenschaftlichen Datensammlung. Bei der Migration werden die retrolektiven Daten Teil des klinischen Registers und müssen entsprechend denselben Qualitätsansprüchen, Plausibilitäts- und Konsistenzkriterien standhalten wie die nativ im Register erhobenen prolektiven Daten. Ist eine historische wissenschaftliche Dokumentation *problembefahet*, woher gegebenenfalls die Motivation für den Aufbau des klinischen Registers herrührt, dann ist vor der Migration eine Plausibilitäts- und Qualitätsprüfung notwendig. In einigen Fällen kann dies sogar eine Quelldatenverifizierung und eine Nachdokumentation erforderlich machen, beispielsweise wenn retrolektive Merkmale nur durch zusätzlich zu transkribierende Informationen auf prolektive Merkmale abgebildet werden können. Die Eigenschaften einer wissenschaftlichen Datensammlung, das verwendete Dokumentationssystem, die Systematik bei Aufbau, Umsetzung und Erfassung und nicht zuletzt die Nutzeranzahl und -schulung bestimmen die Qualität einer wissenschaftlichen Datensammlung. Beispielsweise kann in einer Datenbank die Datenkonsistenz eher sichergestellt werden als in einem Spreadsheet, das von mehreren Benutzern verwendet wird. Daher ist es vor einer Migration notwendig, nicht nur semantische Interoperabilität und eindeutige Identifikation zu beachten, sondern auch die Qualitätsindikatoren der Forschungsdatenquelle. Bei einer Verlinkung sind diese zwar nicht unerheblich, allerdings ist es weniger kritisch und liegt nicht in der Verantwortung des klinischen Registers, wie *gut* eine wissenschaftliche Datensammlung ist. Bei der Migration geht die historische Datenquelle nicht nur informationstechnisch in das klinische Register über, sondern auch inhaltlich und organisatorisch, und daher ist abzuwägen, ob eine einmalige, dauerhafte und vereinte Lösung getrennten und unabhängigen Systemen vorzuziehen ist.

Wie bereits weiter oben erläutert, sind für jegliche Arten der Integration semantische Interoperabilität und eindeutige Identifikation erforderlich. Im Idealfall werden während der Planung des klinischen Registers

Tabelle 12: Gegenüberstellung Integrationsarten bei Dokumentationssystemen der Forschung.

	MIGRATION	VERLINKUNG
Voraussetzung	semantische Interoperabilität eindeutige Identifikation identische Qualitäts- und Konsistenzkriterien	semantische Interoperabilität eindeutige Identifikation
Dauerhaftigkeit	permanent	temporär
Häufigkeit	einmalig	beliebig oft
Ergebnis nach Integration	Forschungsdatenquelle in prolektivem Register enthalten	zusammengeführter Datensatz, Forschungsdatenquelle und Register weiterhin in separaten Systemen

die relevanten Forschungsdatenquellen ermittelt und Merkmale und Instanzen so definiert, dass eine Zusammenführung möglich wird. Sind Datensammlungen sowohl in ihren Merkmalen als auch in ihren Instanzen heterogen, und sind keine Gemeinsamkeiten und Überschneidungen feststellbar, dann ist letztendlich auch eine Integration nicht möglich. Je früher dies festgestellt wird, idealerweise nicht erst bei einer geplanten Zusammenführung und Auswertung, desto eher kann mit geänderten Merkmalsdefinitionen oder zusätzlichen Identifikatoren diesem entgegen gewirkt werden.

Realisierungsaspekte

Die Integration eines prolektiven Dokumentationssystems und einer Forschungsdatenquelle wird in Form eines ETL-Prozesses umgesetzt, bei dem relevante Daten aus Register und Datensammlung extrahiert, in eine einheitliche Struktur transformiert und anschließend in ein Zielsystem geladen werden. Die technische Realisierung des ETL-Prozesses ist dabei unabhängig von der Integrationsart und auch bei der Transkription strukturierter klinischer Informationen anzuwenden (vergleiche Kapitel 4.2.3). Der einzige Unterschied zwischen Verlinkung und Migration besteht darin, ob ein neuer, zusammengesetzter Datensatz erzeugt oder ob in einen bestehenden integriert werden soll. Da im erstgenannten Fall die Integration inhaltlich komplexer sein kann, schließlich sind mehrere Quellstrukturen auf eine neue (vorgegebene) Zielstruktur zu überführen, wird der ETL-Prozess in Zusammenhang mit Data Warehouse im nachfolgenden Kapitel beschrieben.

Voraussetzung für die Realisierung eines ETL-Prozesses zwischen zwei Datensammlungen sind semantische Interoperabilität der Merkmale und eindeutige Identifikation der Instanzen. Werden bei einer Integration zusätzliche Instanzen angefügt, dann ist im Vorfeld die semantische

Interoperabilität zu prüfen und festzustellen, ob sich zwei Merkmale in ihren Definitionen, Eigenschaften und Wertebereichen so ähnlich sind, dass sie aufeinander abgebildet werden können. Unterscheiden sich Merkmale inhaltlich voneinander, beispielsweise da sich Klassifikationssysteme geändert haben oder Erhebungen nach unterschiedlichen Richtlinien erfolgen, dann lässt sich zwar gegebenenfalls syntaktische Heterogenität mittels Transformationsregeln überwinden, eine Integration ist aber aufgrund der verschiedenen Bedeutungsstrukturen nicht möglich. Eine Prüfung, die individuell aufwändig und abhängig von der Anzahl der zu integrierenden Merkmale langwierig sein kann, erfordert immer eine direkte Gegenüberstellung der Merkmale sowie Kenntnis der fachlichen Bedeutung und technischen Repräsentation. Werden Integrationsprozesse wiederholt durchgeführt, dann sollte eine redundante Erhebung und Analyse derselben Datensätze möglichst vermieden und auf bereits vorhandene Informationen zurückgegriffen werden. Unterstützen können dabei sogenannte MDMS, mit deren Hilfe semantische und syntaktische Eigenschaften von Merkmalen erfasst, bewertet und für den Transformationsschritt geeignet ausgegeben werden. Aufgrund der Bedeutung eines MDMS für sämtliche Integrationsarten wird es in Kapitel 4.2.6 gesondert beschrieben.

Die eindeutige Identifikation von Instanzen wird immer dann relevant, wenn in den Dokumentationssystemen der Versorgung und Forschung unterschiedliche Informationen zu denselben Patienten erfasst werden, beispielsweise klinische Daten im EPA-System, administrative Daten in einem PMS oder Studiendaten in einem SDMS. Sind übergreifende Auswertungen vorgesehen, dann müssen vor der Merkmalsintegration die Instanzen in den einzelnen Datensammlungen eindeutig identifiziert werden, sodass eine Verknüpfung über Systemgrenzen hinweg möglich wird. Während semantische Interoperabilität auch bei ähnlichen und aufeinander abbildbaren Merkmalen erfüllt werden kann, ist bei Instanzen eine restriktive und exakte Übereinstimmung notwendig. Häufig wird dies über einen sogenannten *Identifikator* erreicht, d. h. ein bestimmtes Merkmal, das eine Instanz eindeutig kennzeichnet. Identifikatoren sind in der Regel Nummern oder Codes und werden idealerweise systemübergreifend einheitlich und konsistent genutzt. Wurden allerdings innerhalb einer klinisch-wissenschaftlichen Versorgungseinrichtung unterschiedliche Patienten-IDs, studieninterne Nummern oder Pseudonyme etabliert und eine eindeutige Zuordnung ist nicht über ein einzelnes Merkmal möglich, dann sind zusätzliche Identifikationsmerkmale notwendig, um denselben Patienten als solchen zu erkennen. Unter der Annahme, dass die datenschutzrechtlichen Voraussetzungen erfüllt und in den Datensammlungen dieselben Merkmale vorhanden sind, kann eine eindeutige Identifikation auch über die Kombination mehrerer Merkmale erfolgen. Neben naheliegenden Merkmalen wie *Nachname* oder *Geburtsdatum*

sind auch seltene klinische Merkmale wie das *Diagnose-*, *Therapie-* oder *Einschlussdatum* geeignet. Das dabei verwendete Prinzip basiert darauf, redundant erfasste Daten zur Identifikation der Instanzen zu verwenden. Ein manueller Abgleich mehrerer Datensätze ist sowohl zeitaufwändig als auch fehleranfällig und wird durch unterschiedliche Richtlinien, Schreibweisen und Erfassungsfehler zusätzlich erschwert. Zweckmäßiger sind daher automatisierte sogenannte *Record Linkage Verfahren*, die über spezielle Algorithmen Ähnlichkeitsmaße zwischen den Merkmalen zweier Instanzen berechnen und dadurch die Zusammenführung mehrerer Datensammlungen unterstützen. Record Linkage ist häufig als eigenständige Anwendung umgesetzt, oder kann mit Hilfe theoretischer Veröffentlichungen selbstständig implementiert werden (Camelot Consulting 2012; Christen 2012; Jurczyk et al. 2011). Die Verfahren sind in der Regel mehrstufig aufgebaut und sehen für jedes zu vergleichende Merkmal einen eigenen Algorithmus mit individueller Parametrisierungsmöglichkeit vor. Die Schwierigkeit für den Benutzer besteht darin, aus einer Vielzahl an Algorithmen auszuwählen und diese so zu kombinieren, dass dieselben Instanzen in zwei Datensammlungen mit hoher Sensitivität und Spezifität identifiziert werden. Zudem problematisch ist, dass sich die verfügbaren Record Linkage Werkzeuge nur unzureichend in den ETL-Integrationsprozess einreihen und meistens manuelle Zwischenschritte vor der Merkmalstransformation erfordern, sodass Eigenimplementierungen, die individuelle Integrationsprozesse unterstützen, vorzuziehen sind.

Die Umsetzung eines Record Linkage Frameworks wurde mit *Talend Open Studio for Data Integration* implementiert und für zwei Integrationsprozesse verwendet (vergleiche auch Kapitel 5.3 und 5.4) (Löpprich et al. 2014a).

4.2.5 Data Warehouse

Ein Data Warehouse ist in der Referenz-IT-Architektur als themenorientiertes, integriertes und beständiges System zur Entscheidungsunterstützung vorgesehen. Dadurch kann es ergänzend zu den operativen Dokumentationssystemen der Forschung und Versorgung eingesetzt werden und eine globale Sicht auf heterogene und verteilte Datensammlungen ermöglichen.

Im Unterschied zu operativen Dokumentationssystemen sind die in einem Data Warehouse vorhandenen Daten nicht durch Benutzeraktionen veränderbar, sondern werden nur über einen Integrationsprozess aktualisiert. Der Zugriff auf die Daten im Data Warehouse ist damit rein lesend und entspricht dem Stand der Dokumentation zum Zeitpunkt der Integration. Dadurch wird sichergestellt, dass keine Inkonsistenz zwischen

operativen Dokumentationssystemen und integriertem Datenbestand entsteht und Integration immer nur in eine Richtung notwendig ist.

Eine weitere Eigenschaft, die sich aus dem Vorgehensmodell zur systematischen Planung eines klinischen Registers ableitet, ist, dass nicht alle Merkmale aller zu integrierenden Datensammlungen übernommen werden, sondern nur die, die für bestimmte Analysen, Fragestellungen und Entscheidungsfindungen relevant sind. Entsprechend werden beim Integrationsprozess Merkmale aggregiert, klassifiziert, neu berechnet oder weggelassen. Beispielsweise kann ein im Dokumentationssystem als taggenau erfasstes *Geburtsdatum* für verschiedene Auswertungen als *Alter* oder *Altersklasse* besser geeignet sein und in dieser Form in das Data Warehouse integriert werden.

Der Prozess der Informationsintegration erfolgt nach mehrstufigem Prinzip zur Sicherstellung technischer, syntaktischer und semantischer Interoperabilität und eindeutiger Identifikation von Instanzen. Ohne Integration stellt ein Data Warehouse lediglich ein Software- und Datenbanksystem dar. Auf der anderen Seite kann Integration aber auch ohne Data Warehouse notwendig sein und einen integrierten Datenbestand liefern, der mit Hilfe von Statistik- und Auswertungssoftware weiterverarbeitet werden kann. Ein Data Warehouse bietet aber gegenüber einer direkten Analyse weitere Vorteile wie eine webbasierte und einheitliche Darstellung der Daten. Zudem sind Benutzeroberflächen in der Regel einfach bedienbar und erlauben insbesondere wissenschaftlich tätigen Ärzten und Mitarbeitern einen schnellen und direkten Zugriff auf die Daten. Informationsintegration, bei der Daten in eine Zielstruktur transformiert und in eine zentrale Datenbasis übertragen werden, ist meistens ein schwieriger und zeitaufwändiger Prozess und wird umso komplizierter je mehr Datenquellen einbezogen und berücksichtigt werden.

Damit ein Data Warehouse sinnvoll und nachhaltig eingesetzt werden kann, ist es notwendig, dass der Integrationsprozess wiederholt ausgeführt und die Daten regelmäßig aus den operativen Systemen integriert werden. Liegen keine aktuellen Daten im Data Warehouse vor, dann führt dies unter Umständen zu falschen Schlussfolgerungen oder schränkt die Nutzerakzeptanz ein. Da ein Data Warehouse innerhalb eines klinischen Registers selten denselben Stellenwert wie das prolektive Dokumentationssystem haben sollte, ist ein möglichst automatisierter Integrationsprozess mit kurzen Aktualisierungsintervallen und ohne manuelle Zwischenschritte oder zusätzlichen Aufwand empfehlenswert. Die technische Realisierung der Informationsintegration wird als ETL-Prozess bezeichnet und teilt sich in die drei Schritte: Extraktion, Transformation und Laden. Aspekte, die die Planung und Realisierung eines Data Warehouse und des ETL-Prozesses betreffen, werden im Folgenden beschrieben.

Planungs- und Realisierungsaspekte

Ein Data Warehouse stellt neben dem prolektiven Dokumentationssystem des klinischen Registers ein weiteres Softwaresystem dar, dessen Hauptnutzergruppe aber nicht die medizinischen Dokumentare sind, sondern vielmehr die wissenschaftlich tätigen Ärzte und Mitarbeiter. Entsprechend ist bei Auswahl und Parametrisierung auf einfache und intuitive Bedienbarkeit zu achten. Als geeignete OpenSource Systeme sind dabei *izb2* und *tranSMART* zu nennen (Athey et al. 2013; Murphy et al. 2007). Die Umsetzung eines Data Warehouse sollte innerhalb eines klinischen Registers mehr als Service- und Zusatzleistung betrachtet werden und weniger als notwendiges System. Typischerweise ist das prolektive Dokumentationssystem primär zu realisieren, gefolgt von den Integrationsprozessen der Versorgung und Forschung. Relevanz, unterschiedliche Benutzergruppen, Nutzungshäufigkeit und Verwendungszweck legen zudem eine Trennung von prolektivem Dokumentationssystem und Data Warehouse auf Serverebene nahe. Zweckmäßig ist es daher, das Data Warehouse auf einem eigenen virtuellen oder physischen Server zu betreiben und die operativen Systeme des klinischen Registers davon getrennt zu halten, sodass Ressourcen, Daten- und Ausfallsicherung überwiegend für das EDC-System geplant und eingesetzt werden können. Sind Integrationsprozesse etabliert, dann ist ein neuer integrierter Datensatz einfach zu erzeugen und ein gegebenenfalls ausgefallenes Data Warehouse schnell wieder in Betrieb zu nehmen, ohne jeglichen Datenverlust.

Der ETL-Prozess der Informationsintegration hat das Ziel, Daten aus mehreren heterogenen Datensammlungen zu extrahieren, zu transformieren und in die Zieldatenbank des Data Warehouse zu laden. Alternativ besteht auch die Möglichkeit, den integrierten Datensatz ohne Data Warehouse auszugeben und manuell weiterzuverarbeiten oder auszuwerten. Der umfangreichste Schritt der Informationsintegration ist die Transformation, bei der syntaktische und semantische Heterogenität überwunden und eine eindeutige Identifikation hergestellt wird. Unterstützende Werkzeuge und Methoden dabei sind ein IDMS zur eindeutigen Identifikation der Instanzen sowie ein MDMS zur Überwindung semantischer Konflikte. IDMS und MDMS sind eigenständige Systeme, die manuell gepflegt und fortlaufend betrieben werden müssen, um einen nahtlosen Integrationsprozess zu ermöglichen. Wurde beispielsweise ein Patient im Dokumentationssystem des klinischen Registers oder in einer wissenschaftlichen Datensammlung neu erfasst und die Identifikationsnummern im IDMS nicht aktualisiert, dann ist auch die eindeutige Identifikation und Integration dieses Patienten nicht möglich. Ähnliches gilt auch beim MDMS, wenn neu zu integrierende Datensammlungen erschlossen und neue Merkmale eingegeben oder bestehende Definitionen angepasst werden müssen.

Liegen keine aktuellen Informationen im IDMS und MDMS vor, dann kann im Regelfall auch die Informationsintegration nicht funktionieren. OpenSource Werkzeuge zur Realisierung des ETL-Prozesses sind unter anderem *Pentaho Data Integration* oder *Talend Open Studio for Data Integration* (Pentaho Corporation 2016; Talend 2016).

Extraktion

Bei der Extraktion der Daten aus verschiedenen Quellsystemen ist zunächst die Frage des technischen Zugriffs zu klären: Werden Daten über Softwareschnittstellen der Dokumentationssysteme geliefert, besteht ein direkter Zugriff auf das DBS oder werden Dateien (idealerweise in Standarddateiformaten) übermittelt? Manuelle Benutzeraktionen wie das Verschieben von Dateien auf Netzlaufwerke sollten möglichst vermieden und automatisierte Verfahren präferiert werden. Weiter ist wichtig, dass die Extraktion immer nach derselben zuvor definierten Methode verläuft und dass das Ergebnis dieselben Merkmale und Spaltenreihenfolge enthält. Integrationsprozesse werden gewöhnlich für eine feste Menge an Merkmalen und Datenquellen angelegt und sind anzupassen, sobald sich Merkmale und/oder Quellen ändern, d. h. variable ETL-Prozesse, die mit beliebigen Datensammlungen umgehen können, sind sehr selten und sollten auch in einem ersten Realisierungsprozess nicht angestrebt werden. Die Art des Zugriffs legt häufig auch implizit die Aktualisierungsintervalle fest und ob Daten synchron und damit bei jeder Änderung in den Quellsystemen, oder asynchron in regelmäßigen Abständen, übermittelt werden. Während bei Softwareschnittstellen die Übertragung in der Regel synchron erfolgt, werden Exporte manuell, zeit- oder ereignisgesteuert erstellt. Nicht bei jeder Änderung ist ein Integrationsprozess notwendig, sondern häufig erst dann relevant, wenn dieselbe neue Instanz in mehreren Datensammlungen angelegt wurde. Daten können zwar synchron oder in Form eines kompletten Datenbestandes aus einer Datenquelle extrahiert werden, sind dann allerdings so lange zwischenspeichern, bis die Aktualisierungen aller zu integrierender Quellen eingegangen sind. Durch die Extraktion wird technische Heterogenität zwischen den verschiedenen Hard- und Softwaresystemen überwunden und die Daten auf einem für die Transformation zugänglichen System oder Netzlaufwerk abgelegt.

Transformation

Der Transformationsschritt erfordert nacheinander das Überwinden syntaktischer und semantischer Heterogenität und die eindeutige Identifikation der Instanzen. Abhängig von der Anzahl zu integrierender Datenquellen kann dies unterschiedlich komplex sein. Im einfachsten Fall werden nur die Daten aus einem prolektiven Dokumentationssystem abgefragt und in ein anderes Zielschema überführt, beispielsweise in das

eines spezifischen Auswertungswerkzeugs oder eines Data Warehouse. Falls sich Quell- und Zielschema voneinander unterscheiden, sind verschiedene Formate, Datentypen und Modelle aufeinander abzubilden, was in der Regel individuelle und situationsbedingte Analysen der Datensammlungen und Zielstruktur notwendig macht. Um syntaktische Heterogenität möglichst gering zu halten, ist es daher sinnvoll, bereits bei der Extraktion der Daten auf Standarddateiformate und ähnliche Datenstrukturen zu achten sowie gleichbleibende Systemeigenschaften bei erneuten Integrationsprozessen vorauszusetzen. Besteht Zugriff auf die Daten, dann sind unterschiedliche Zeichencodierungen und Datenformate über Transformations- und Ersetzungsregeln aufzulösen, wie die Umwandlung eines amerikanischen in ein deutsches Datumsformat oder von Dezimalpunkten in -kommas.

Da ein Integrationsprozess auch immer ein Harmonisierungsprozess zwischen einer oder mehreren Datenquellen und einem Zielsystem darstellt, ist ein MDMS notwendig, um semantische Interoperabilität zwischen Merkmalen festzustellen und Konflikte zu lösen. Wird ein MDMS verwendet und sind semantische Unterschiede und Gemeinsamkeiten über konzeptionelle Beziehungen zwischen den einzelnen Merkmalen abgebildet, dann wurden Benennungs-, Wertebereichs- und Skalierungskonflikte bereits vorher betrachtet und nur noch Abbildungs- oder Transformationsregeln sind zu definieren. Die Transformation mehrerer Quellstrukturen auf eine neue, zusammengeführte Struktur ist dabei weniger komplex, als wenn eine bestimmte Struktur des Zielsystems vorgegeben wird. So ist beispielsweise die verwendete Wissensrepräsentationssprache, Taxonomie oder Ontologie eines Data Warehouse produktabhängig und erfordert weitere individuelle Anpassungen der semantisch interoperablen Merkmale auf das Zielsystem. Die eindeutige Identifikation einer Instanz in mehreren Datensammlungen setzt entweder eine einheitliche Identifikationsnummer oder ein IDMS voraus. Alternativ besteht auch die Möglichkeit über Record Linkage Verfahren die Identität einer Instanz zu ermitteln und eine Zuordnung herzustellen. Dazu ist allerdings eine Teilmenge an identischen Merkmalen in den Datensammlungen notwendig, über die dieselbe Instanz eindeutig identifiziert und in einem zusammengeführten Datensatz als solche dargestellt werden kann.

Das Ergebnis der Transformation ist ein einzelner zusammengesetzter Datensatz, der entweder ausgegeben oder in das DBS eines Data Warehouse geladen wird.

Laden

Beim Laden wird der integrierte Datensatz über System- oder Datenbankschnittstellen in das DBS des Data Warehouse überführt. Um lesende Zugriffe während der Transaktion zu vermeiden, sollten die

Importfunktionen des DBS genutzt oder das Laden außerhalb der Arbeitszeit durchgeführt werden.

Wurde ein ETL-Prozess einmalig definiert, bleiben Datenquellen und Rahmenbedingungen unverändert und sind IDMS und MDMS aktuell, dann kann die Informationsintegration wiederholt ablaufen und einen zusammengeführten Datensatz automatisiert in einem Data Warehouse bereitstellen.

4.2.6 Metadatenmanagementsystem

In klinisch-wissenschaftlichen Versorgungseinrichtungen existieren häufig viele verschiedene Datensammlungen, die über Jahre hinweg entstanden sind und sich unterschiedlichen wissenschaftlichen Fragestellungen widmen. Aufgrund inhaltlicher, zeitlicher oder räumlicher Distanz können sich diese Datensammlungen in Definition, Syntax und Systematik deutlich voneinander unterscheiden. Der Umgang mit verschiedenen heterogenen Datensätzen und das Sicherstellen semantischer und syntaktischer Interoperabilität stellt eine große Herausforderung bei der Integration mit den Dokumentationssystemen der Versorgung und Forschung dar. Begründet und motiviert wird eine Integration häufig durch übergreifende Auswertungen und einen erhofften zusätzlichen Informationsgewinn, der prinzipiell auf zwei Arten zustande kommen kann: entweder zu bestehenden Instanzen existieren weitere Merkmale, oder zu bestehenden Merkmalen existieren weitere Instanzen. Im erstgenannten Fall ist die eindeutige Zuordnung und Identifikation derselben Instanz in verschiedenen Systemen notwendig. Ist diese möglich und sind die Merkmalsmengen disjunkt, dann sollte eine Integration beider Systeme inhaltlich verhältnismäßig einfach realisierbar sein. Überschneiden sich jedoch Merkmale, bzw. sind in verschiedenen Systemen dieselben oder ähnliche Merkmale enthalten, dann ist zunächst die Interoperabilität der Systeme festzustellen. Wird beispielsweise in einem System das *Geschlecht* mit *männlich* / *weiblich* und in einem anderen mit *female* / *male* / *unknown* codiert, dann sind vor der Datenintegration zunächst die Eigenschaften der einzelnen Merkmale zu prüfen und gegebenenfalls eindeutige Regeln für eine konsistente Zusammenführung festzulegen. Analog zu der Identifikation einer Instanz, die bei jeder Integration notwendig wird, müssen auch für jedes nicht-disjunkte Merkmal die Eigenschaften vorher bestimmt und Transformationsregeln definiert werden. Während man bei Instanzen einrichtungswert eindeutige Identifikationsnummern oder IDMS nutzt, bietet sich bei Merkmalen ein sogenanntes *Metadatenmanagementsystem* (MDMS) an. Metadaten sind Informationen über Syntax und Semantik einzelner Merkmale und erlauben, verwaltet in einem MDMS, die Darstellung von Inkompatibilitäten zwischen Sys-

temen sowie die Machbarkeitsbewertung geplanter Auswertungen und Integrationen.

Sind zwei Systeme zu integrieren, bzw. zwei Datensammlungen zusammenzuführen, dann müssen Metadaten immer zunächst ermittelt und Abbildungsregeln festgelegt werden. Eine systematische Erhebung in Form eines MDMS ist allerdings nicht immer erforderlich, beispielsweise wenn die Integration, wie bei der Migration, nur einmalig vorgesehen ist. Zusätzlich ist bei einem MDMS zu beachten, dass es ein von dem prolektiven Dokumentationssystem des klinischen Registers unabhängiges System ist und keine Informationen zu wissenschaftlichen Fragestellungen liefert, sondern lediglich Merkmalsbeschreibungen, die den Integrationsprozess unterstützen und beschleunigen. Der Inhalt des MDMS ist aber ebenso aktiv zu pflegen und fortlaufend zu dokumentieren wie die wissenschaftlichen Merkmale, da andernfalls der praktische Nutzen schnell nachlässt. Auch für die Transkription, also die Integration mit den Dokumentationssystemen der Versorgung, lässt sich ein MDMS zur Abbildung semantischer Beziehungen zwischen klinischen Informationen und zu dokumentierenden Merkmalen einsetzen. Am sinnvollsten erscheint es aber vorwiegend bei der Verlinkung des prolektiven Dokumentationssystems des klinischen Registers mit wissenschaftlichen Datensammlungen aufgrund der wiederholenden Integration zum Zweck gemeinsamer Auswertungen. Daher wurde in der Referenz-IT-Architektur in Abbildung 1 das MDMS auch nur an dieser Stelle vorgesehen, kann aber prinzipiell auch bei den anderen Integrationsarten eingesetzt werden.

Für die Entwicklung eines MDMS kann der Standard *ISO / IEC 11179, Information Technology – Metadata registries (MDR)* genutzt werden. Da dieser äußerst umfangreich und komplex ist, werden in diesem Leitfaden nur die wichtigsten Punkte dargestellt. Diese beziehen sich insbesondere auf das grundlegende logische Datenmodell, das im Folgenden so aufbereitet wurde, dass eine praktische Umsetzung in ein physisches Datenmodell und damit in die Systemarchitektur eines MDMS möglich wird.

Planungs- und Realisierungsaspekte

Bei der Umsetzung eines MDMS ist in einem ersten Schritt festzulegen, welche Metadaten pro Merkmal erhoben werden sollen. Diese sollten mindestens die sein, die in dem Registerprotokoll zur Beschreibung der Merkmale verwendet wurden, d. h. die *Bezeichnung* auf der Oberfläche, der *Variablennamen*, der *Datentyp* und die *Wertemenge*. Anschließend gilt es für jedes Merkmal aus den zu integrierenden Datensammlungen diese Metadaten zusammenzutragen. Wurden Systeme und Datensammlungen systematisch geplant, dann können hierfür die jeweiligen Merkmalskataloge (vergleiche Registerprotokoll Protokollpunkt 5.1.2) verwendet

werden. Mit Hilfe der Merkmalskataloge verschiedener Datensammlungen lassen sich häufig bereits unterschiedliche Benennung semantisch ähnlicher Merkmale identifizieren. Die Abbildung semantischer Beziehungen zwischen Merkmalen unterschiedlicher Datensammlungen in einem MDMS erfolgt über ein sogenanntes *Konzept*. In einem physischen Datenmodell kann die Verbindung zwischen abstraktem Konzept und konkretem Merkmal als 1:N Beziehung dargestellt werden: einem Konzept sind mehrere Merkmale, meist ein Merkmal pro Datensammlung, zugeordnet und jedes Merkmal wird durch genau ein Konzept beschrieben.

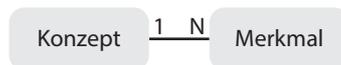


Abbildung 10: Beziehung zwischen abstraktem Konzept und konkretem Merkmal zur Beschreibung von Metadaten.

In der Praxis zeigt sich, dass die Beziehung zwischen Konzept und Merkmal nicht immer 1:N sein muss, sondern davon auch abweichen kann. Diese Fälle sollten explizit gemacht werden und in Form eines *abgeleiteten Merkmals* dargestellt werden. Näheres dazu wird in Tabelle 13 beschrieben.

Wird das MDMS so realisiert, dass nicht nur Merkmale semantisch in Relation zueinander gesetzt werden können, sondern auch deren Wertebereiche, dann ist dies über eine N:1 Beziehung möglich. Damit kann ein konkreter *Wertebereich* mehreren Merkmalen zugeordnet werden und ein Merkmal genau einen Wertebereich haben. Typische Wertebereiche verschiedener Merkmale sind beispielsweise *ja / nein / unbekannt* oder *vorhanden / nicht vorhanden*.

Tabelle 13: Ableitung von Merkmalen zur Vermeidung unterschiedlicher konzeptioneller Bedeutungen.

Abhängig von einer bestimmten Fragestellung oder Auswertung sind häufig mehrere Konzepte für ein konkretes Merkmal definierbar. Laut *ISO / IEC 11179* ist aber ein Merkmal immer eindeutig einem Konzept zuzuordnen und über Transformationsregeln abgeleitete Merkmale sind als neues Merkmal zu werten. Entsprechend wird dem Konzept *Alter* nicht das konkrete Merkmal *Geburtsdatum* sondern das abgeleitete Merkmal *heutiges Datum minus Geburtsdatum* zugewiesen. Dadurch wird die Konsistenz der semantischen Beziehung sichergestellt und vermieden, dass ein Merkmal mehrere konzeptionelle Bedeutungen haben kann.



Abbildung 11: Beziehung zwischen konkretem Merkmal und Wertebereich zur Beschreibung von Metadaten.

Ein Wertebereich entspricht immer der konkreten Repräsentation eines Merkmals und kann sich von Datensammlung zu Datensammlung unterscheiden. Um die Bedeutungsgleichheit zwischen konkreten Merkmalsausprägungen sicherzustellen, wird das physische Datenmodell um eine weitere Ebene *Wert* mit einer M:N Beziehung zum Wertebereich ergänzt. Ein Wertebereich besteht üblicherweise aus mehreren Werten und derselbe semantische Wert kann mehreren Wertebereichen zugeordnet sein.

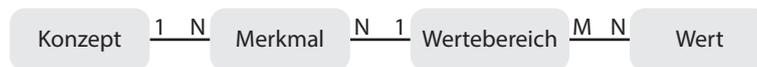


Abbildung 12: Hergeleitetes physisches Metadatenmodell eines Metadatenmanagementsystems.

Durch die M:N Beziehung zwischen Wertebereich und Wert wird erreicht, dass zusätzlich zu der Semantik von Merkmalen (über das Konzept) auch die Semantik der Merkmalsausprägungen (über die Werte) beschrieben werden kann. Diese lassen sich damit direkt in einem ETL-Prozess als Abbildungsregel für den Transformationsschritt nutzen.

Abbildung 13 zeigt das physische Datenmodell eines MDMS ergänzend an einem Beispiel und stellt grafisch dar, wie mehrere semantisch ähnliche Merkmale mit einem abstrakten Konzept und mehrere semantisch ähnliche Merkmalsausprägungen mit einem abstrakten Wert in Beziehung zueinander gesetzt werden können.

Die Eingabe und Pflege von Metadaten in einem MDMS erfordert ein hohes Maß an fachlichem und inhaltlichem Verständnis. Nur dadurch kann sichergestellt werden, dass semantische Beziehungen und Unterschiede zuverlässig erkannt und festgehalten werden. Ist dies möglich, dann bietet ein MDMS eine große Unterstützung bei der Integration und Zusammenführung mehrerer Dokumentationssysteme.

Zusätzlich zu den oben genannten Eigenschaften lassen sich weitere Informationen zu Merkmalen erheben. Relevant wird dies insbesondere dann, wenn mit Hilfe eines MDMS nicht nur semantische Abhängigkeiten analysiert, sondern auch die Machbarkeit einer geplanten Auswertung oder die Quantität und Qualität einer Integration vor der eigentlichen Ausführung festgestellt werden soll. Metadaten zur Vollständigkeit, dem

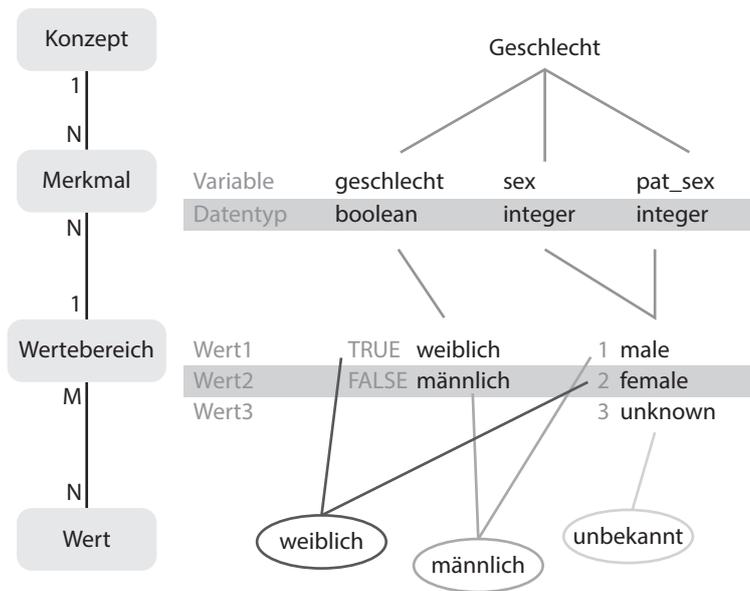


Abbildung 13: Beispiel zum physischen Metadatenmodell. Die drei konkreten Merkmale *geschlecht*, *sex* und *pat_sex* entsprechen allen demselben abstrakten Konzept *Geschlecht*. Die Merkmale *sex* und *pat_sex* haben denselben Wertebereich. Die konkreten Werte *weiblich* und *female* sind über den Wert *weiblich* semantisch identisch, die konkreten Werte *männlich* und *male* über den Wert *männlich*.

Dokumentationszeitraum oder der Art der Qualitätssicherung liefern weitere detaillierte Angaben zu den verschiedenen Datensammlungen und ermöglichen einen strukturierten Überblick über das Vorliegen relevanter Daten und derer Merkmale.

Für die praktische Umsetzung eines MDMS eignen sich vorwiegend relationale Datenbanksysteme mit denen das Metadatenmodell implementiert werden kann. Abhängig davon, ob ein Ein- oder Mehrbenutzerbetrieb realisiert werden soll, ist eine Webanwendung einer Desktopanwendung vorzuziehen und eine Benutzeroberfläche zu erstellen. Intuitive Benutzeroberflächen sind aufgrund der verschiedenen Beziehungstypen (1:N, N:1 und M:N) nicht trivial und häufig aufwändig in ihrer Entwicklung. Sind Benutzer mit der Eingabe und Bedienung einer Datenbankanwendung vertraut, dann ist eine separate Benutzeroberfläche nicht zwingend erforderlich. Ein relationales DBS bietet sich auch daher an, da Transformations- und Abbildungsregeln für spezifische Integrationen über Abfragen einfach bestimmt werden können. Zweckmäßig ist auch eine Kommunikationsmöglichkeit mit dem Data Warehouse bzw. dem ETL-Prozess vorzusehen und die im MDMS gespeicherten

semantischen Beziehungen automatisch für die Harmonisierung und Transformation verschiedener Datensammlungen zu nutzen.

Die praktische Umsetzung eines MDMS auf Basis des hergeleiteten physischen Metadatenmodells wurde für das klinische Myelomregister des Universitätsklinikums Heidelberg erstellt und ist im Kapitel 5.5 beschrieben.

4.2.7 Hinweise zur Anwendung

Die Referenz-IT-Architektur wurde als abstrahiertes Modell der logischen Werkzeugebene eines klinischen Registers und seiner Anwendungssysteme entwickelt und erlaubt bei einer konkreten Umsetzung unterschiedliche Implementierungsvarianten. Die wesentlich zu treffenden Entscheidungen, die zu einer konkreten Architektur führen, sind mit Verweis auf die einzelnen Anwendungssysteme und Komponenten in Abbildung 14 dargestellt. Weitere Entscheidungen, die für jedes Anwendungssystem relevant sind, werden ausführlich in den jeweiligen Kapiteln dargelegt und beschrieben und sind in der Tabelle 14 zusammenfassend aufgeführt.

Tabelle 14: Zu treffende Entscheidungen bei konkreter Umsetzung der Referenz-IT-Architektur.

PROLEKTIVES DOKUMENTATIONSSYSTEM

- Ist ein Dokumentationssystem zur prolektiven Erfassung von Daten notwendig oder sind ausschließlich retrolektive Daten zu integrieren?
 - Soll ein Dokumentationssystem ausgewählt und parametrisiert werden oder ist eine Eigenentwicklung möglich?
 - Ist ein dynamisches oder ein statisches Dokumentationsmodell mit dem Dokumentationssystem abzubilden?
 - Ist ein Studiendatenmanagementsystem als Dokumentationssystem geeignet oder sind zusätzliche registerspezifische Funktionen relevant?
 - Sind verschiedene Dokumentationssysteme zur Trennung identifizierender oder klinischer Daten notwendig?
 - Sind Prozesse zur Dokumentation identifizierender und klinischer Daten unabhängig voneinander oder ist eine gemeinsame Sicht auf identifizierende und klinische Daten relevant?
-

Tabelle 14: (Fortsetzung)

INTEGRATION MIT SYSTEMEN DER VERSORGUNG

- Sind die zu integrierenden Informationen digital im Krankenhausinformationssystem verfügbar?
- Sind weitere Informationen von konventionellen Dokumenten relevant?
- Sind Daten aus dem rechnerunterstützten Teil des Krankenhausinformationssystems exportierbar oder kann über Schnittstellen darauf zugegriffen werden?
- Ist eine vollständig manuelle Transkription leistbar oder sind automatisierte Werkzeuge zur Unterstützung zu entwerfen?

INTEGRATION MIT SYSTEMEN DER FORSCHUNG

- Soll eine existierende Datensammlung durch das prolektive Dokumentationssystem des klinischen Registers ersetzt werden?
- Überschneiden sich zu integrierende Datensammlungen in ihren Merkmalen oder in ihrem Patientenkollektiv?
- Sind in den zu integrierenden Datensammlungen dieselben Identifikationsmerkmale vorhanden oder ist ein Record Linkage notwendig?
- Sind in den zu integrierenden Datensammlungen dieselben Merkmalsdefinitionen vorhanden oder sind semantische Beziehungen sowie Abbildungs- und Transformationsregeln notwendig?

DATA WAREHOUSE

- Sollen mehrere Datensätze für wissenschaftliche Analysen integriert werden?
- Soll der integrierte Datenbestand in einem Data Warehouse verfügbar gemacht oder in Form eines auswertbaren Datensatzes bereitgestellt werden?
- Wie häufig und in welchen Abständen soll die Integration aus den operativen Dokumentationssystemen erfolgen?

METADATENMANAGEMENTSYSTEM

- Ist die Integration heterogener Datensammlungen häufig für verschiedene Fragestellungen notwendig?
 - Existieren Merkmalsbeschreibungen der Datensammlungen oder sind Metadaten über die Inhalte zu bestimmen?
 - Können Metadaten regelmäßig geprüft und konsistent zu den Datensammlungen gehalten werden?
-

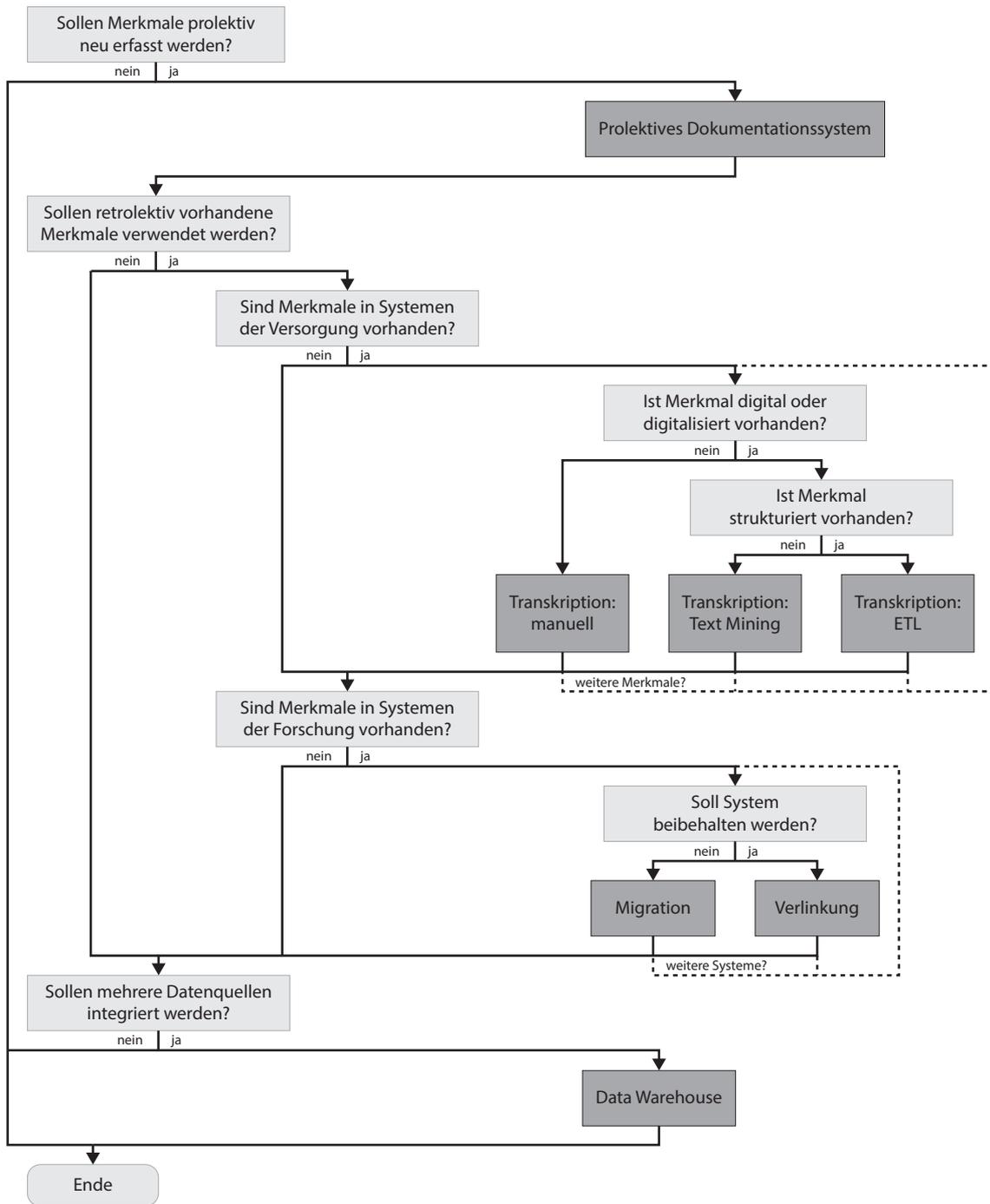


Abbildung 14: Darstellung wesentlicher Entscheidungen zur Umsetzung der Referenz-IT-Architektur. Eine gestrichelte Linie stellt dabei die Möglichkeit einer Wiederholung dar, wenn beispielsweise mehrere Merkmale oder Systeme vorhanden sind.

ANWENDUNG DES LEITFADENS AM BEISPIEL DES KLINISCHEN MYELOMREGISTERS

Der im vorangegangenen Kapitel entwickelte Leitfaden, bestehend aus Vorgehensmodell und Referenz-IT-Architektur, wurde im Rahmen eines Kooperationsprojekts zwischen der *Sektion Multiples Myelom (Medizinische Klinik V des Universitätsklinikums Heidelberg, Nationales Centrum für Tumorerkrankungen)* und der *Sektion Medizinische Informatik (Institut für Medizinische Biometrie und Informatik)* zur Planung und Aufbau eines neu einzuführenden klinischen Registers angewendet. Ziel war die Umsetzung einer informationsverarbeitenden und speichernden Infrastruktur für Patientendaten mit der hämatologisch-onkologischen Erkrankung *Multiples Myelom*.

In diesem Kapitel wird zunächst das Registerprotokoll des klinischen Myelomregisters beschrieben. Für die Realisierung der IT-Infrastruktur gemäß der Referenz-IT-Architektur wurden einige allgemein anwendbare Methoden und Werkzeuge entwickelt. Daher werden in den folgenden Kapiteln die Ergebnisse der Bewertung prolektiver Dokumentationssysteme, die Realisierung eines Text Mining Werkzeugs zur Integration unstrukturierter und eines ETL-Prozesses zur Integration strukturierter Informationen sowie ein System zur Speicherung und Verwaltung von Metadaten vorgestellt.

5.1 REGISTERPROTOKOLL KLINISCHES MYELOMREGISTER

Gemäß dem Vorgehensmodell von Kapitel 4.1 wurde das Registerprotokoll für das klinische Myelomregister des Universitätsklinikums Heidelberg in enger Abstimmung mit dem Datenmanagement und Ärzten abgeleitet. In dieser Arbeit wird aus Gründen der Vertraulichkeit und Transparenz bei manchen Protokollpunkten eine verkürzte Fassung beschrieben.

1. Einleitung

1.1 Beteiligte Einrichtung(en)

Die am klinischen Myelomregister beteiligten dokumentierenden Einrichtungen sind:

1. *Sektion Multiples Myelom*
Medizinische Klinik V des Universitätsklinikums Heidelberg
Nationales Centrum für Tumorerkrankungen

2. *GMMG-Studiengruppe*
Universitätsklinikum Heidelberg
3. *Institut für Humangenetik*
Universitätsklinikum Heidelberg

Neben den dokumentierenden Einrichtungen ist die *Sektion Medizinische Informatik* des *Instituts für Medizinische Biometrie und Informatik (IMBI)* der *Universität Heidelberg* an Planung, Aufbau und Betrieb des klinischen Registers beteiligt. Die statistische Auswertung der Registerdaten erfolgt sektionsintern oder durch die *Abteilung Biostatistik* des *Deutschen Krebsforschungszentrums (DKFZ)*.

1.2 Beteiligte Interessengruppen und Personen

Die am klinischen Myelomregister beteiligten Interessengruppen und jeweils verantwortlichen Personen sind:

1. Registerleitung: H.G.
2. Wissenschaftliches Datenmanagement der Sek. Multiples Myelom
Ansprechpartner: M.P.
3. Wissenschaftliche Dokumentation der Sektion Multiples Myelom
Ansprechpartner: K.S.-L.
4. Administrative Verwaltung der Sektion Multiples Myelom
Ansprechpartner: A.A.
5. GMMG-Studiengruppe
Ansprechpartner: U.B.
6. Institut für Humangenetik des Universitätsklinikums Heidelberg
Ansprechpartner: A.J.
7. Sektion Medizinische Informatik
Ansprechpartner: M.L.
8. Abteilung Biostatistik des Deutschen Krebsforschungszentrums
Ansprechpartner: A.K.-S.
9. Wissenschaftlich tätige Ärzte der Sektion Multiples Myelom

1.3 Gegenstand und Motivation

In der Sektion Multiples Myelom werden für Myelompatienten außerhalb klinischer Interventionsstudien im Rahmen einer wissenschaftlichen Dokumentation seit 1992 individuelle Krankheits- und Therapieverläufe dokumentiert und auch nach Abschluss der Behandlung weiter verfolgt. Das Ziel ist die Bearbeitung und Auswertung wissenschaftlicher Fragestellungen zum Multiplen Myelom. Dazu werden klinische Daten mit hämatologischen, molekularbiologischen und studienspezifischen Daten korreliert und als Konferenz- oder Zeitschriftenbeiträge publiziert.

Da die myelomspezifischen Daten über Jahrzehnte hinweg in unterschiedlichen Systemen erfasst und gespeichert wurden, liegen sie aktuell in heterogenen Dateiformaten und Definitionen vor. Wissenschaftliche Anfragen und übersichtliche Darstellungen des Behandlungsgeschehens erfordern eine manuelle Zusammenführung mehrerer Dateien, die zeit- und ressourcenaufwändig sind. Zudem haben sich Behandlungs- und Therapiemethoden im Verlauf geändert und erfordern eine inhaltliche und technische Anpassungen der strukturierten Erfassung.

2 Definition der Registerziele

2.1 Problemstellung

Aus der oben beschriebenen Motivation ergeben sich im Einzelnen nachstehende Probleme:

PROBLEM 1 Die Behandlung eines Myelompatienten erfolgt in der Routineversorgung individuell und lokale, systemische und supportive Therapiemaßnahmen werden variabel kombiniert und patientenspezifisch angewendet. Der Krankheitsverlauf ist häufig rezidivierend und Behandlungen werden wiederholend innerhalb eines Zeitraums von mehreren Jahren durchgeführt. Eine wissenschaftliche Dokumentation und standardisierte Erfassung aktueller Behandlungs- und Therapiemethoden erfordert inhaltliche, strukturelle und technische Anpassungen.

PROBLEM 2 Die Behandlung eines Myelompatienten erfolgt interdisziplinär mit Beteiligung verschiedener Kliniken, Institute, Studiengruppen und niedergelassenen Fachärzten. Die dabei entstehenden Daten sind heterogen bezüglich ihrer Herkunft, Struktur und Qualität. Strukturierte automatisierte Zugriffe auf existierende Datensammlungen für übergreifende wissenschaftliche Auswertungen sind bislang nicht etabliert und werden daher zeitaufwändig manuell realisiert.

PROBLEM 3 Die vorhandenen historischen Datensammlungen sind heterogen hinsichtlich Dateiformat, Struktur und Qualität und erfordern zeitaufwändige manuelle Bearbeitungsschritte für übergreifende Auswertungen.

2.2 Zielsetzung

ZIEL 1 Ziel ist die standardisierte Erfassung des gesamten Behandlungs- und Therapieverlaufs eines individuellen Myelompatienten zur genauen Abbildung der Versorgungsroutine und übersichtlichen Darstellung von Krankheitsverläufen. Patienten sind von der Diagnose des Multiplen Myeloms oder den Vorstufen bis zum Versterben

zu dokumentieren. Informationen zum Behandlungsverlauf und zur Progredienz sind in internen oder externen Arztbriefen im EPA-System verfügbar. Angaben zum Überleben sind alternativ über staatliche Ämter zu ermitteln.

ZIEL 2 Ziel ist die Beantwortung wissenschaftlicher Fragestellungen und die Weitergabe standardisiert erfasster myelomspezifischer Daten für übergreifende wissenschaftliche Auswertungen und zur Korrelationen mit molekularbiologischen und studienspezifischen Datenquellen, die den Vergleich unterschiedlicher Therapieschemata und Wirksamkeitsanalysen erlauben. Dabei ist redundante Datenerfassung zu vermeiden und ein Zugriff auf existierende Datenquellen zu realisieren.

ZIEL 3 Ziel ist die Aufbereitung und Nutzung der historischen Datensammlungen für neue wissenschaftliche Fragestellungen und die Weiterdokumentation offener Patienten und Fälle im Rahmen der standardisierten Erfassung von Ziel 1.

3 Bewertung der Ausgangssituation

3.1 Aktuelle Versorgungssituation und Kenntnisstand

Das Multiple Myelom ist eine chronisch rezidivierende Krebserkrankung der Plasmazellen im Knochenmark. Durch eine Vermehrung von Plasmazellen steigt die produzierte Menge eines bestimmten Immunglobulins. Die erhöhte Konzentration verursacht auf Dauer Symptome wie Organ- und Knochenschädigungen und führt letztendlich zum Tod. Jährlich erkranken etwa 3000 Männer und 2700 Frauen in Deutschland. Die Vorstufen des Multiplen Myeloms sind die Monoklonale Gammopathie unbestimmter Signifikanz (MGUS), das solitäre Plasmozytom (solitPC) und das asymptomatische Myelom. Die Therapie des Multiplen Myelom ist äußerst komplex, setzt sich aus mehreren Therapieabschnitten zusammen und erstreckt sich in der Regel über mehrere Jahre. Induktionstherapie, Mobilisierung, Hochdosis-Chemotherapie mit anschließender autologer Blutstammzelltransplantation und Erhaltungstherapie gelten aktuell als Standardbehandlung bei Patienten unter 70 Jahren. Alternativ kann auch ein konventionelles Therapieschema angewendet werden, allerdings mit niedriger Remissionsrate. Tritt nach der Primärtherapie keine Remission ein, werden ein oder mehrere Rezidivtherapien eingeleitet. Wurden genügend autologe Blutstammzellen in der Primärtherapie gesammelt, sind bis zu drei Transplantationen möglich. Anschließend wird in der Regel konventionell oder palliativ weiterbehandelt.

Der Schwerpunkt der Sektion Multiples Myelom am Universitätsklinikum Heidelberg liegt auf der Hochdosis-Chemotherapie mit autologer Stammzelltransplantation und auf Therapien mit neuen Wirkstoffen,

innovativen Substanzen und supportiven Maßnahmen. Stellenwert, optimaler Zeitpunkt und nebenwirkungsarme Therapieformen werden im Rahmen klinischer Studien der GMMG-Studiengruppe wissenschaftlich untersucht.

3.2 Vorhandene Datenquellen

Informationen zur Routinebehandlung eines Myelompatienten stammen aus Arztbriefen und Originalbefunden, die im EPA-System des Universitätsklinikums Heidelberg gespeichert sind. Arztbriefe werden von Fachärzten der Sektion Multiples Myelom erstellt und sind elektronisch über Benutzerschnittstellen zugänglich. Befunde stammen meist aus hämatologischen, immunchemischen oder bildgebenden Verfahren und sind Bestandteil der Arztbriefe. Alle zu einem Behandlungsfall verfügbaren Laborbefunde können als CSV-Datei exportiert werden. Innerhalb des EPA-Systems angelegte Arztbriefe sind in der Regel gleich aufgebaut und einheitlich gegliedert und enthalten überwiegend Freitext. Weiter relevant sind Arztbriefe niedergelassener Ärzte mit Informationen zu extern durchgeführten Behandlungen und Untersuchungen. Externe Arztbriefe werden von einem Scandienstleister des Universitätsklinikums Heidelberg digitalisiert und sind im Archivdokumentationssystem über Benutzerschnittstellen verfügbar. Informationen über Patientenaufnahme und Untersuchungstermine sind über Ambulanzlisten bestimmbar. Diese liegen strukturiert vor und sind über Benutzerschnittstellen einsehbar, können aber nicht exportiert werden.

Daten klinischer Interventionsstudien der GMMG-Studiengruppe sind über das GMMG-Studiensekretariat anzufragen. Bisher wurden mehrere Studien zur Primär- und Rezidivtherapie des Multiplen Myeloms durchgeführt, die teilweise noch aktiv sind. GMMG-Studien sind multizentrisch und enthalten neben Patienten, die am Universitätsklinikum Heidelberg rekrutiert und behandelt wurden / werden, auch Patienten externer Häuser.

Im Rahmen eines Kooperationsprojekts zwischen der Sektion Multiples Myelom und dem Institut für Humangenetik des Universitätsklinikums Heidelberg werden Informationen zur molekularen Zytogenetik in einer eigenen Forschungsdatenbank gespeichert. Alle am Universitätsklinikum Heidelberg behandelten Myelompatienten werden dabei mittels Fluoreszenz-in-situ-Hybridisierung (FISH) untersucht und relevante Chromosomenaberrationen erfasst. Die Dokumentation erfolgt strukturiert merkmalsbasiert und ein elektronischer Zugriff auf das Datenbanksystem ist möglich.

Historische Datensammlungen sind als Excel-Datei und in einem Datenbanksystem verfügbar. Die Excel-Datei wurde zwischen 1992 und 2009 überwiegend von den wissenschaftlich tätigen Ärzten der Sektion Multiples Myelom gepflegt. Das Datenbanksystem wurde von 2009 bis

2014 ausschließlich von medizinischen Dokumentaren genutzt. Der Beobachtungsgegenstand in der Excel-Datei ist die Stammzelltransplantation und wurde in dem Datenbanksystem auf die gesamte Myelomerkrankung eines Patienten mit Vorstufen und zusätzlichen Therapiemöglichkeiten erweitert. Die Patienten in beiden Systemen sind disjunkt, die Merkmale überlappen sich teilweise, können aber aufeinander abgebildet werden. Bei Abschluss 2009 bzw. 2014 enthielt die Excel-Datei ca. 1000, das Datenbanksystem ca. 1200 Patienten. Eine Weiternutzung der in beiden historischen Datensammlungen verfügbaren myelomspezifischen Daten für Auswertungen des Myelomregisters ist vorgesehen.

3.3 Infrastrukturelle Anbindung

Die Sektion Multiples Myelom ist Teil der Medizinischen Klinik V des Universitätsklinikums Heidelberg und des Nationalen Centrum für Tumorerkrankungen. Als Einrichtungen zur infrastrukturellen Anbindung wurden nachfolgende Institute identifiziert:

1. Zentrum für Informations- und Medizintechnik (ZIM) des Universitätsklinikums Heidelberg
Administration und Betrieb des elektronischen Teil des Krankenhausinformationssystems und der zugehörigen Anwendungssysteme; technische Systemverwaltung und Serverhosting; Bereitstellung virtueller Server für Mitarbeiter des Universitätsklinikums
2. Koordinierungszentrum für Klinische Studien (KKS) des Universitätsklinikums Heidelberg
Projektmanagement, Beratung und Durchführung klinischer Studien; Daten- und Qualitätsmanagement inklusive Medizincontrolling und Audits; Administration und Betrieb von Studiendatenmanagementsystemen
3. Institut für Medizinische Biometrie und Informatik (IMBI) der Universität Heidelberg
Biometrisch-statistische und informationstechnische Methodenentwicklung und Beratung wissenschaftlicher Projekte

3.4 Möglichkeiten und Grenzen des Registers

Die Machbarkeit der in Kapitel 2.3 aufgeführten Ziele wird unter den beschriebenen Rahmenbedingungen als realistisch eingeschätzt. Die gesamten klinischen Informationen liegen in elektronischem Format vor und sind mindestens über Benutzerschnittstellen zugänglich. Da molekularbiologische und studienspezifische Datenquellen im Verantwortungsbereich der Sektion Multiples Myelom liegen, sind wenig organisatorische und technische Hindernisse zu erwarten. Heterogene Datenstrukturen und Definitionen sowie individuelle Behandlungs- und Therapieverläufe stellen

eine Herausforderung bei Zusammenführung und Auswertung dar, die die Implementierung spezifischer Anwendungssysteme erfordern.

Planung, Aufbau und initialer Betrieb des Myelomregisters sind bis Mai 2017 projektfianziert. Eine Weiterfinanzierung wird über Drittmittel angestrebt.

4 Spezifikation des Registers

4.1 Fragestellungen

Die für das Myelomregister spezifizierten wissenschaftlichen Fragestellungen beziehen sich auf die Gesamtheit an Informationen, die sich aus der standardisierten Erfassung individueller Verläufe (Ziel 1), der Zusammenführung mit molekularbiologischen und studienspezifischen Datenquellen (Ziel 2) und der Integration historischer Datensammlungen (Ziel 3) ergeben. Eine Auswahl repräsentativer Fragestellungen ist nachfolgend aufgeführt:

1. Welches sind typische Verläufe für den Progress der Myelomkrankung und wie häufig sind diese?
2. Welche Therapieverfahren und Verläufe verlängern das Gesamtüberleben oder begünstigen eine Langzeitremission?
3. Welche molekularbiologischen, genetischen oder histologischen Faktoren beeinflussen die Entstehung eines Multiplen Myeloms und seiner Vorstufen?
4. Welche Komorbiditäten sind mit einer ungünstigen Prognose beim Multiplen Myelom assoziiert?
5. Unterscheiden sich Patienten der Routinebehandlung von Studienpatienten hinsichtlich des Gesamtüberlebens, der progressionsfreien Zeit und des Remissionsstatus?

4.2 Patientenkollektiv

4.2.1 Zielkollektive

Für das Myelomregister werden zwei Zielkollektive definiert:

1. Patienten mit diagnostiziertem MGUS, solitärem Plasmozytom (solitPC), asymptomatischem Myelom, Plasmazell-Leukämie (PC-L) oder POEMS-Syndrom
 - keine Studienteilnahme
 - keine Therapieintention
2. Patienten mit diagnostiziertem symptomatischem Myelom
 - keine Studienteilnahme
 - Therapieintention zur Hochdosis-Chemotherapie mit autologer Blutstammzelltransplantation
 - Alter bei Therapieintention ≤ 75 Jahre

4.2.2 Auswertungskollektive

Für die Auswertungskollektive werden folgende Einschlusskriterien vorgegeben:

1. Patienten mit diagnostiziertem MGUS, solitärem Plasmozytom (solitPC), asymptomatischem Myelom, Plasmazell-Leukämie (PC-L) oder POEMS-Syndrom
 - Vorstellig in der Sektion Multiples Myelom des Universitätsklinikums Heidelberg mit Erstvorstellungsdatum ≥ 2009 (inklusive Patienten aus historischem Datenbanksystem)
 - schriftliches Einverständnis
2. Patienten mit diagnostiziertem symptomatischem Myelom
 - Vorstellig in der Sektion Multiples Myelom des Universitätsklinikums Heidelberg mit Erstvorstellungsdatum ≥ 1992 (inklusive Patienten aus historischem Datenbanksystem und historischer Excel-Datei)
 - Therapiedurchführung der Hochdosis-Chemotherapie mit erster autologer Blutstammzelltransplantation durch die Sektion Multiples Myelom des Universitätsklinikums Heidelberg
 - schriftliches Einverständnis

4.3 Registermethodik

4.3.1 Behandlungs- und Versorgungsmethodik

Die Erstvorstellung eines Myelompatienten in der Sektion Multiples Myelom erfolgt in der Regel zur Komplettierung der Diagnostik oder zur Einholung einer Zweitmeinung und wird meistens von einem niedergelassenen Haus- oder Facharzt initiiert. Die Diagnostik eines Multiplen Myeloms, seiner Vor-, Zwischen- und Endstufen, die Stadieneinteilung und Feststellen der Behandlungspflichtigkeit sieht spezifische hämatologische und immunchemische Laborparameter, Knochenmark- und bildgebende Analysen vor. Diagnose und diagnostische Verfahren werden bei Krankheitsprogression und/oder Stadienwechsel wiederholt. Abhängig von der initialen Diagnose, dem Patientenalter und seinem allgemeinen Gesundheitszustand sind lokale Vortherapien, wie Operationen, Bestrahlungen oder systemische Chemotherapien notwendig. Werden Patienten für eine Hochdosis-Chemotherapie mit autologer Blutstammzelltransplantation vorbereitet, werden nacheinander Induktionstherapie mit mehreren Therapiezyklen, Mobilisierung und Stammzellsammlung, Therapieintensivierung und Transplantation, Konsolidierungs- und Erhaltungstherapie durchgeführt. Therapieabschnitte werden patientenspezifisch variiert und bei Unverträglichkeit gewechselt oder abgebrochen. Wird nach der Stammzelltransplantation keine Remission erreicht, können Hochdosistherapie

und Transplantation wiederholt durchgeführt werden. Hochdosistherapie und Stammzelltransplantation können stationär, die anderen Therapieabschnitte teilweise auch bei niedergelassenen Hämatologen erfolgen. Therapieansprechen und Reduktion der Tumorlast werden nach jedem Therapieabschnitt über dieselben Verfahren wie bei der Diagnostik bestimmt. Ergebnisse extern durchgeführter Behandlungen werden schriftlich übermittelt. Wird nach der Primärtherapie eine komplette Remission erreicht, finden Kontrollen und Untersuchungen mindestens halbjährig statt. Kommt es zum Krankheitsprogress wird eine Rezidivtherapie eingeleitet und Induktionstherapie, sowie Therapieintensivierung und erneute Transplantation werden durchgeführt. Sind keine weiteren Transplantate von der Stammzellsammlung verfügbar, wird nach konventionellen Chemotherapieschemata weiterbehandelt. Überleben und Erkrankungsstatus werden intern bestimmt oder von extern angefragt, Überlebensstatus und Sterbedatum können zusätzlich von staatlichen Ämtern ermittelt werden.

4.3.2 Erhebungsmethodik

Die Erhebung relevanter Informationen zum Behandlungs- und Therapieverlauf eines Myelompatienten erfolgt während der Routinebehandlung durch Fachärzte der Sektion Multiples Myelom in Form von Freitexten in elektronischen Arztbriefen. Labor-, Radiologie- und Knochenmarkbefunde sind ebenfalls im EPA-System des Universitätsklinikums Heidelberg als Freitext verfügbar und werden bei der Arztbriefschreibung vom Facharzt eingefügt und bewertet. Externe Befunde und Arztbriefe liegen zunächst papierbasiert, später digitalisiert im Archivdokumentationssystem vor. Die Erhebung dient primär der Patientenversorgung und ist nur eingeschränkt standardisiert möglich.

Überlebens- und Erkrankungsstatus sind von niedergelassenen Fachärzten oder staatlichen Ämtern im Rahmen der Langzeitnachbeobachtung von Myelompatienten aktiv zu erheben. Anfragen erfolgen schriftlich und sehen keine standardisierte Rückmeldung vor.

4.3.3 Integrationsmethodik

Intern angelegte und extern digitalisierte Arztbriefe sind im EPA-System über Benutzerschnittstellen zugänglich und können manuell in das klinische Myelomregister übertragen werden. Die manuelle Transkription erfolgt täglich durch medizinische Dokumentare. Laboruntersuchungen sind im CSV-Dateiformat exportierbar und können über einen ETL-Prozess transkribiert werden. Der Exportvorgang muss manuell und pro Fall initiiert werden und ist täglich vorgesehen.

Die Integration mit molekularbiologischen Daten der Zytogenetik-Forschungsdatenquelle kann über eine Verlinkung realisiert werden. Patienten sind durch die im Universitätsklinikum Heidelberg genutzte

Patientenidentifikationsnummer eindeutig identifizierbar. Alternativ enthält die Forschungsdatenquelle weitere identifizierende Merkmale, über die ein Record Linkage möglich ist. Verlinkung und Aktualisierung der integrierten Daten erfolgt wöchentlich.

Die Integration mit den studienspezifischen Datenquellen der GMMG-Studien kann über eine Verlinkung realisiert werden. Eine Zusammenführung ist nur für die am Universitätsklinikum Heidelberg rekrutierten und behandelten Patienten möglich. Patienten sind in den einzelnen Datenquellen durch eine studieninterne Nummer identifizierbar. Weitere identifizierende Merkmale können über das Studiensekretariat angefragt werden, sind jedoch in den Datenquellen nicht enthalten. Zusammenführungen werden bei Bedarf durchgeführt.

Die Integration der historischen Datenquellen kann für offene und nicht abgeschlossene Patienten und Fälle in Form einer einmaligen Migration erfolgen. Die Daten sind anschließend aus den historischen Systemen zu entfernen. Patienten, die bereits verstorben sind, können über eine Verlinkung integriert werden. Identifikation ist über die Patientenidentifikationsnummer möglich. Da keine Aktualisierung abgeschlossener Fälle und Patienten vorgesehen ist, erfolgt die Verlinkung einmalig.

4.3.4 Standardisierte Erfassungs- und Dokumentationsmethodik

National oder international anerkannte Datensatz- und Merkmalsdefinitionen zur Erhebung und Dokumentation einer Myelomerkrankung existieren nicht. Für einzelne Merkmale sind verschiedene myelomspezifische Klassifikations- und Ordnungssysteme vorgesehen. Eine beispielhafte Auswahl:

- das International Staging System (ISS) der International Myeloma Foundation
- die CRAB-Kriterien der International Myeloma Working Group
- die Response-Kriterien der International Myeloma Working Group
- die Durie-Salmon Stadienklassifikation

4.3.5 Statistische Auswertungsmethodik

Die meisten Fragestellungen lassen sich über Methoden der deskriptiven Statistik auswerten. Gesamtüberleben (Overall Survival; OS) und Progressionsfreies Überleben (Progression-Free Survival; PFS) werden nach der Kaplan-Meier-Methode bestimmt. Genaue Zeitpunkte und Umfang der Auswertungen können vorab nicht eindeutig festgelegt werden. Ein statistischer Analyseplan wird daher jeweils unmittelbar vor einer geplanten Auswertung verfasst.

4.3.6 Präsentationsmethodik und Formulare

Es werden exemplarisch zwei Präsentationslayouts zur Beantwortung der Fragestellungen 1. und 2. dargestellt. Die angegebenen Zahlen sind fiktiv und dienen ausschließlich der Veranschaulichung.

Welches sind typische Verläufe für den Progress der Myelomerkrankung und wie häufig sind diese?

	MGUS	solitPC	asym. Myelom	sym. Myelom	PC-L	POEMS
MGUS	17 %	6 %	12 %	61 %	3 %	1 %
solitPC	-	... %	... %	... %	... %	... %
asym. Myelom	-	-	... %	... %	... %	... %
sym. Myelom	-	-	-	... %	... %	... %
PC-L	-	-	-	-	... %	... %
POEMS	-	-	-	-	-	... %

Abbildung 15: Präsentationslayout zur Beantwortung der Fragestellung nach typischen Verläufen der Myelomerkrankung. Vertikal wird die Vorstufe abgebildet, horizontal der Progress. Die Angaben sind wie folgt zu interpretieren: 17 % der Patienten mit MGUS hatten keinen Progress, 6 % der Patienten sind in das Stadium solitäres Plasmozytom übergegangen usw.

Welche Therapieverfahren verlängern das Gesamtüberleben?

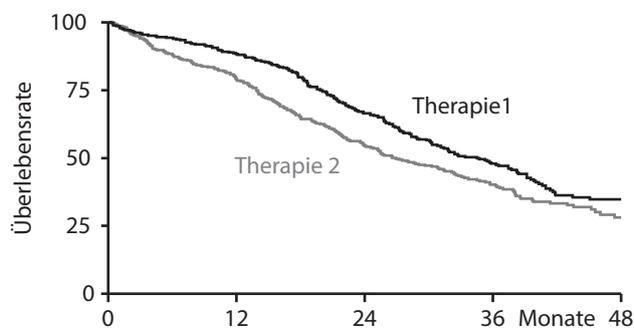


Abbildung 16: Präsentationslayout zur Beantwortung der Fragestellung nach dem Gesamtüberleben. Auf der vertikalen Achse ist die Überlebensrate in % und auf der horizontalen Achse die Anzahl der Monate abgebildet. Die Angaben sind wie folgt zu interpretieren: Nach 24 Monaten haben ca. 70 % der Patienten mit Therapie 1 überlebt, verglichen zu 50 % der Patienten mit Therapie 2.

5 Entwurf des Registers

5.1 Datenhaltung

5.1.1 Entwurf des Registerschemas

Für das Registerschema des klinischen Myelomregisters wurden folgende Domänen definiert: Patient, Arzt, Diagnose, einzelne Therapieabschnitte (Lokaltherapie, Chemotherapie, Mobilisierung und Intensivierung), einzelne Befunde (Hämatologie, Immunologie, Bildgebung und Knochenmark), Therapieansprechen und Überlebensstatus. Jeder Patient wird durch mindestens eine Diagnose charakterisiert. Die Erfassung externer Ärzte ist für die Erhebung externer Therapieabschnitte, Therapieansprechen und dem Überlebensstatus notwendig. Jeder Patient kann keinen, einen oder mehrere niedergelassene Ärzte haben und ein niedergelassener Arzt kann einen oder mehrere Patienten behandeln. Die zeitliche Abfolge von Diagnosen, Therapieabschnitten, Befunden und Therapieansprechen ist variabel und kann wiederholt bei einem Patienten auftreten. Die Erfassung von Behandlungs- und Therapieverläufen wird als dynamisches Dokumentationsmodell umgesetzt.

Zur Trennung identifizierender und klinischer Daten werden Angaben zu Patienten und niedergelassenen Ärzten in einem separaten Identitätsdatenmanagementsystem (IDMS) gespeichert. Die Zuordnung zu den klinischen Daten erfolgt über ein Pseudonym.

5.1.2 Merkmalskatalog

In Tabelle 16 werden exemplarisch einige repräsentative Merkmale der Domänen *Patient*, *Diagnose* und *Hämatologie* genannt. Insgesamt wurden 402 Merkmale spezifiziert.

5.2 Datenmanagement

5.2.1 Organisation der Datenerfassung

Die Datenerfassung im klinischen Myelomregister erfolgt, sofern nicht anders spezifiziert, manuell durch einen medizinischen Dokumentar auf Grundlage interner und externer Arztbriefe, die papierbasiert oder digitalisiert im Archivdokumentationssystem oder elektronisch im EPA-System über Benutzerschnittstellen zugänglich sind.

Eine neue Instanz der Domäne *Patient* wird initial mit *Patienten-Nr.*, *Nachname*, *Vorname*, *Geschlecht* und *Geburtsdatum* im IDMS angelegt, sobald auf der Ambulanzliste der kommenden Woche ein Untersuchungstermin für einen neuen Patienten angelegt wurde. Der Abgleich der Patienten auf der Ambulanzliste und im IDMS erfolgt manuell. Der behandelnde Arzt am Universitätsklinikum Heidelberg klärt den Patienten im Anschluss an den Untersuchungstermin über das klinische Myelomregister auf und händigt die Einverständniserklärung zur Unterschrift aus. Eingang und Einverständnis werden im IDMS dokumentiert und im KDAT-System wird die Domäne *Überlebensstatus* mit aktuellem

Tabelle 15: Auszug aus dem Merkmalskatalog des klinischen Myelomregisters. Es sind die drei Domänen Patient, Diagnose und Hämatologie mit repräsentativen Merkmalen dargestellt.

NR.	VARIABLE	BEZEICHNUNG	DATENTYP	WERTEMENGE	DATENQUELLE
Patient					
1	idPatient	Patienten-Nr.	Zahl	-	EPA-System
2	idPseudonym	Pseudonym	Zahl	{100000-499999}	wird generiert
3	lastname	Nachname	Text	-	EPA-System
4	firstname	Vorname	Text	-	EPA-System
5	sex	Geschlecht	Einfachauswahl	0, unbekannt 1, männlich 2, weiblich	EPA-System
Diagnose					
1	idPseudonym	Pseudonym	Zahl	{100000-499999}	IDAT-System
2	diag_date	Diagnosedatum	Datum	-	Arztbrief, Diagnose
3	diag_source	Quelle	Einfachauswahl	1, intern 2, extern	Arztbrief
4	diag	Diagnose	Einfachauswahl	1, symp. Myelom 2, asym. Myelom 3, MGUS ...	Arztbrief, Diagnose
5	diag_staging	Stadium nach Durie-Salmon	Einfachauswahl	1, IA 2, IB 3, IIA ...	Arztbrief, Diagnose
Hämatologie					
1	idPseudonym	Pseudonym	Zahl	{100000-499999}	IDAT-System
2	hema_date	Datum Hämatologie	Datum	-	Arztbrief, Laborbefunde
3	hema_source	Quelle	Einfachauswahl	1, intern 2, extern	Arztbrief
4	ca_available	Calcium verfügbar	Einfachauswahl	0, nein 1, ja	Arztbrief, Laborbefunde
5	ca	Calcium Wert	Zahl	-	Arztbrief, Laborbefunde
6	ca_unit	Calcium Einheit	Einfachauswahl	0, unbekannt 1, mmol/l 2, mg/dl	Arztbrief, Laborbefunde

LastInfo-Datum angelegt. Die Zuordnung zum Patienten erfolgt über ein Pseudonym. Ist ein Patient mit aktuellem Untersuchungstermin bereits im IDMS enthalten, werden im EPA-System die vorhandenen internen und externen Arztbriefe geprüft und Instanzen der jeweiligen Domänen nach Bedarf und enthaltenen Informationen in den Arztbriefen im KDAT-System angelegt. Die Zuordnung zum Patienten, die zeitliche Reihenfolge und der Quellverweis werden über die entsprechenden Merkmale für alle Domänen dokumentiert. Klinische Informationen zum Krankheits- und Behandlungsverlauf werden im KDAT-System erfasst, Informationen zum externen Arzt als neue Instanz der Domäne *Arzt*.

5.2.2 Organisation des Follow-Up

Patienten, die über einen längeren Zeitraum keinen neuen Untersuchungstermin am Universitätsklinikum Heidelberg hatten und von denen keine neuen Informationen in Form externer Arztbriefe vorliegen, werden nachverfolgt. Dazu wird halbjährig eine Abfrage im KDAT-System durchgeführt und nach Patienten mit dem Überlebensstatus *lebend* und dem *LastInfo-Datum* ≤ 6 Monate für die Diagnose *symptomatisches Myelom* bzw. *LastInfo-Datum* ≤ 12 Monate für alle anderen Diagnosen gefiltert. Die Ergebnisliste wird mit Hilfe des IDMS aufgelöst und die behandelnden niedergelassenen Ärzte werden identifiziert und schriftlich kontaktiert. Erfolgt keine Rückmeldung wird das Einwohnermeldeamt am Wohnort des Patienten angefragt. Rückmeldungen werden von medizinischen Dokumentaren im KDAT-System erfasst.

5.2.3 Rechnerunterstützte Datenerfassung

Die Anforderungen an das elektronische Datenerfassungssystem des klinischen Myelomregisters sind:

- Zugriff über einen Webbrowser innerhalb des Universitätsklinikums Heidelberg
- Plausibilitätsregeln zur Sicherung der Datenqualität
- Report- und Berichtsfunktion u. a. zur Organisation des Follow-Up
- Benutzerauthentifizierung über Benutzername und Passwort
- Zugriff auf das Datenbanksystem für Integrationsprozesse
- Trennung IDAT und KDAT
- keine Eigenentwicklung

Die informationstechnischen Werkzeuge werden im Protokollpunkt 6.3 beschrieben. Die Installation erfolgte auf einem physischen Server im Zentrum für Informations- und Medizintechnik (ZIM) des Universitätsklinikums Heidelberg.

5.2.4 Rechnerunterstützte Kommunikation

Der rechnergestützte Zugriff auf die im EPA-System angelegten Laborbefunde ist über eine Exportfunktion möglich. Der Export aller Laborparameter eines Behandlungsfalls wird manuell initiiert und erfolgt als CSV-Datei auf ein Netzlaufwerk. Weiterverarbeitung und Speicherung im KDAT-System werden über ein selbstentwickeltes ETL-Werkzeug realisiert.

Der Zugriff auf die Zytogenetik-Forschungsdatenquelle erfolgt über TCP / IP auf das Datenbanksystem. Studienspezifische Daten sind beim GMMG-Studiensekretariat anzufragen und werden in variierenden Dateiformaten abhängig von den jeweiligen Studiendokumentationssystemen zur Verfügung gestellt. Abgeschlossene Fälle und Patienten in den historischen Datenquellen sind als Excel-Datei direkt, bei dem Datenbanksystem über TCP / IP verfügbar.

Daten des klinischen Myelomregisters werden bei Bedarf über Berichtsfunktionen oder über das Datenbanksystem abgefragt und die Weitergabe an die wissenschaftlich tätigen Ärzte der Sektion Multiples Myelom erfolgt als Exportdatei.

5.2.5 Sicherung der Datenqualität

Ein Großteil der Plausibilitätsprüfungen wird direkt im klinischen Myelomregister umgesetzt und erfolgt während der Dateneingabe. Zusätzliche Qualitätskontrollen werden mit spezifischen Statistikwerkzeugen wie SAS durchgeführt und Fehler manuell von medizinischen Dokumentaren in dem KDAT-System korrigiert. Vor jeder Datenweitergabe zur wissenschaftlichen Auswertung ist eine separate Qualitätskontrolle des exportierten Datensatzes vorgesehen.

Es werden exemplarisch einige repräsentative Plausibilitätsprüfungen des Validierungsplans genannt:

Tabelle 16: Auszug aus den Plausibilitätsprüfungen des klinischen Myelomregisters.

NR.	DOMÄNE(N)	VARIABLE(N)	REGEL	FEHLERMELDUNG
1	Diagnose	diag_date	Muss kleiner als heutiges Datum sein.	Diagnosedatum liegt in der Zukunft. Bitte prüfen.
2	Hämatologie	ca	Muss zwischen 1.7 und 2.6 mmol/l liegen.	Hinweis: Calciumwert liegt außerhalb des Wertebereichs. Bitte prüfen.
3	Diagnose Hämatologie	diag_staging crea	Bei Durie-Salmon-Stadium IA, IIA und IIIA muss crea < 2.0 mg/dl sein.	Durie-Salmon-Stadium stimmt nicht mit Kreatininwert überein. Bitte prüfen.

5.2.6 Entwurf der Erfassungsformulare

Für jede Domäne ist jeweils ein eigenes Erhebungsformular vorgesehen. Exemplarisch werden zwei Erhebungsformulare dargestellt:

Patient	
Patienten-Nr.	<input type="text" value="1001"/>
Pseudonym	<input type="text" value="307516"/> <input type="button" value="Generieren"/>
Stammdaten	
Titel	<input type="text" value="keiner"/>
Nachname	<input type="text" value="Mustermann"/>
Vorname	<input type="text" value="Max"/>
Geburtsdatum	<input type="text" value="01.01.1980"/> 36
Geschlecht	<input type="text" value="männlich"/>

Chemotherapie	
Quelle	<input type="text" value="intern"/>
Startdatum	<input type="text" value="01.01.2016"/>
Enddatum	<input type="text" value="05.07.2016"/>
Dauer	6 Monat(e)
Zyklen	<input type="text" value="2"/>
Zyklus 1	
	<input type="text" value="Bortezomib"/>
-	<input type="text" value="Adriamycin"/>
-	<input type="text" value="sonstiges"/> <input type="text" value="Prednison"/>
<input type="button" value="+"/>	
Zyklus 2	
	<input type="text" value="Melphalan"/>
-	<input type="text" value="sonstiges"/> <input type="text" value="Rituximab"/>
<input type="button" value="+"/>	

Abbildung 17: Erhebungsformulare der Domänen *Patient* und *Chemotherapie* mit den entsprechenden zu erfassenden Merkmalen.

5.3 Sicherungskonzept

5.3.1 Datenschutz

Die Daten werden auf einem physischen Server im ZIM des Universitätsklinikums Heidelberg im Auftrag der Sektion Multiples Myelom gespeichert. Das ZIM trägt dafür Sorge, dass die Daten vor unberechtigtem Zugriff geschützt sind. Zugriffe aus dem Intranet des Universitätsklinikums liegen im Verantwortungsbereich der Sektion Multiples Myelom und werden über Benutzerauthentifizierung sichergestellt.

Identifizierende und klinische Daten werden in getrennten Datenbanksystemen auf unterschiedlichen virtuellen Servern gespeichert. Der Personenbezug erfolgt über ein Pseudonym, das bei schriftlichem Patienteneinverständnis von den medizinischen Dokumentaren generiert wird. Wird das Patienteneinverständnis widerrufen, ist das Pseudonym im IDMS manuell zu entfernen. Die identifizierenden Informationen *Patienten-Nr.*,

Nachname, Vorname und *Geburtsdatum* bleiben zusammen mit dem Status des Einverständnisses (*Datum, Erteilt / Nicht Erteilt, Widerrufen*) im IDMS gespeichert. Ein positives Ethikvotum zu Datenerhebung, Follow-Up und Nutzung ist vorhanden. Zugriffsberechtigungen auf die jeweiligen Systeme werden restriktiv und individuell vergeben.

Integrierte Daten werden in einem Data Warehouse gespeichert. Personenbezug und Datenexport daraus sind nicht möglich. Zugriffsberechtigungen sind für die wissenschaftlich tätigen Ärzte der Sektion Multiples Myelom vorgesehen.

Die Weitergabe von Daten für wissenschaftliche Auswertungen innerhalb des Universitätsklinikums Heidelberg erfolgt pseudonymisiert, an Stellen außerhalb des Universitätsklinikums anonymisiert. Gehen Datenauswertungs- und Integrationsprojekte über die Zweckbindung des positiven Ethikvotums hinaus, ist es Aufgabe der Sektion Multiples Myelom ergänzende Ethikvoten und/oder Einverständniserklärungen einzuholen.

5.3.2 Datensicherung

Die vom ZIM des Universitätsklinikums Heidelberg angebotenen Maßnahmen zur System- und Datensicherung werden in Anspruch genommen. Zusätzlich sind in regelmäßigen Zeitpunkten manuelle Datensicherungen der Datenbanksysteme durch die Sektion Multiples Myelom auf externen Datenträgern vorgesehen.

6. Betrieb des Registers

6.1 Verantwortliche und Berechtigungen

Folgende Verantwortlichkeiten werden festgelegt:

- Gesamtprojekt: Registerleitung
- Entwicklung und Wartung der informationstechnischen Werkzeuge: Sektion Medizinische Informatik
- Datenerfassung: Dokumentare der jeweiligen dokumentierenden Einrichtung
- Auswertungen: der jeweils Auswertende

6.2 Zeitplanung

Die Zeit- und Projektplanung erfolgte in *Microsoft Project* und wird an dieser Stelle nur verkürzt wiedergegeben:

Entwicklung:	01.08.2013 bis 31.12.2014
Einführung und Tests:	01.01.2015 bis 31.12.2015
Betrieb und Wartung:	01.01.2016 bis auf weiteres
Auswertungen:	01.01.2016 bis auf weiteres

6.3 Informationstechnische Werkzeuge

Die in 5.2.3 spezifizierten Anforderungen an die rechnerunterstützte Datenerfassung werden mit einem physischen Server im ZIM des Universitätsklinikums Heidelberg umgesetzt. Die Serverinstanz CAN₁₀ ist in mehrere virtuelle Server mit unterschiedlichen Anwendungssystemen aufgeteilt (vergleiche Tabelle 17). Als Hypervisor wird *ESXi* von *VMware Global* eingesetzt, als Betriebssystem der virtuellen Server *CentOS*.

Die IT-Infrastruktur basiert auf der Trennung identifizierender und klinischer Daten. Auswahl und Bewertung der verschiedenen Software-systeme erfolgte systematisch und ist als externes Dokument abgelegt (vergleiche Kapitel 5.2).

Das KDAT-System ist mit dem OpenSource EDC-System *OpenClinica* und dem Datenbanksystem *PostgreSQL* auf CAN₀₁ realisiert, das IDMS mit dem OpenSource EDC-System *REDCap* und dem Datenbanksystem *MySQL* auf CAN₀₂. Eine gemeinsame Sicht auf IDAT und KDAT wird durch Integration beider Systeme über den Portalserver *Liferay* auf CAN₀₅ umgesetzt. Ein EPL-Portlet kapselt ausgewählte identifizierende Daten der *MySQL* Datenbank in Listenform, ein KDAT-Portlet die gesamte *OpenClinica* Oberfläche. Ein Instanzaufruf in *OpenClinica* erfolgt durch Aufruf einer URL mit dem Pseudonym des Patienten mit folgendem Aufbau: <https://openclinica.krz.uni-heidelberg.de/OpenClinica/ViewStudySubject?id=idPseudonym>. Änderung und Eingabe identifizierender Daten ist exklusiv in *REDCap* (CAN₀₂), klinischer Daten entweder direkt in *OpenClinica* (CAN₀₁) oder über das KDAT-Portlet von CAN₀₅ vorgesehen.

Auf CAN₀₆ ist das Datenbanksystem *MySQL* installiert und speichert unter anderem die molekularbiologischen Daten des Kooperationsprojekts zwischen der Sektion Multiples Myelom und dem Institut für Humangenetik des Universitätsklinikums Heidelberg. Neue projektspezifische Datenbanken sind auf CAN₀₆ vorgesehen.

Tabelle 17: Serververteilung des klinischen Myelomregisters.

HOSTNAME	VIRTUELLER SERVER	ANWENDUNGSSYSTEM
openclinica.krz.uni-heidelberg.de	CAN ₀₁	KDAT-EDC-System
redcap.krz.uni-heidelberg.de	CAN ₀₂	Identitätsdatenmanagementsystem
mdms.krz.uni-heidelberg.de	CAN ₀₃	Metadatenmanagementsystem
i2b2.krz.uni-heidelberg.de	CAN ₀₄	Data Warehouse
mmportal.krz.uni-heidelberg.de	CAN ₀₅	Portalserver Liferay
mmprojekte.krz.uni-heidelberg.de	CAN ₀₆	Datenbanksystem

Die Integration klinischer Daten von CANo1, molekularbiologischer Daten von CANo6, der historischen Excel-Datei und dem historischen Datenbanksystem erfolgt auf CANo4. Integrierte Daten werden in dem Data Warehouse *izb2* auf CANo4 gespeichert.

Während der Entwicklungsphase wurde das eigenentwickelte Metadatenmanagementsystem auf Basis des Datenbanksystems *PostgreSQL* auf CANo3 genutzt. Eine Weiternutzung ist momentan nicht vorgesehen.

7 Änderungen gegenüber der letzten Protokollversion
Es handelt sich um die erste Version des Registerprotokolls.

5.2 BEWERTUNG ELEKTRONISCHER DATENERFASSUNGSSOFTWARE

Die Auswahl eines Systems zur elektronischen Erfassung klinischer Informationen im Myelomregister erfolgte gemäß den spezifizierten Anforderungen des Registerprotokolls. Als mögliche Produkte wurden die zwei OpenSource Systeme *OpenClinica* und *REDCap* sowie das kommerzielle System *Onkostar* ausgewählt und systematisch miteinander verglichen. Die Bewertung erfolgte auf Grundlage einer zum damaligen Zeitpunkt aktuellen Version der CIPROS Prüfliste (Lindörfer u. Mansmann 2014), die hinsichtlich der Anforderungen und Eigenschaften des Myelomregisters angepasst und um eigene Kriterien erweitert wurde. Die resultierende Liste der Bewertungskriterien ist im Anhang B enthalten und kann zur Beurteilung weiterer Softwareprodukte und zur Prüfung ihrer Eignung für den Einsatz in klinischen Registern genutzt werden. Ebenfalls im Anhang B ist beispielhaft die Bewertung und direkte Gegenüberstellung der beiden Systeme *OpenClinica* und *Onkostar* aufgeführt.

Die Auswertung der einzelnen Kriterien erfolgte nach einer fünfstufigen Skala mit ++ (Eigenschaft sehr gut erfüllt), +, o (Eigenschaft mittelmäßig erfüllt), - bis - - (Eigenschaft nicht erfüllt oder nicht umsetzbar). Falls eine Eigenschaft nicht getestet werden konnte, wurde dies mit ? gekennzeichnet und, sofern möglich, das erwartete Ergebnis dahinter angegeben. Die Gewichtung wurde gemäß den eigenen Anforderungen mit *sehr wichtig*, *mäßig wichtig* und *unwichtig* vorgenommen.

Der wesentliche Unterschied zwischen den OpenSource Systemen *OpenClinica* und *REDCap* war, dass mit *REDCap* nur die Definition statischer Dokumentationsmodelle möglich war, wohingegen mit *OpenClinica* auch dynamische Dokumentationsmodelle und damit individuelle Krankheits-, Behandlungs- und Therapieverläufe abgebildet werden konnten. *OpenClinica* war zwar komplexer in Installation, Parametrisierung und Anwendung, allerdings dafür auch flexibler bei der Dokumentation versorgungs- und realitätsnaher Verläufe. Gegenüber *Onkostar* unterschieden sich *OpenClinica* und *REDCap* dadurch, dass Nutzung und

Installation auf eigenen Servern möglich war und damit eigene Anpassungen an Oberfläche und Datenmodell umgesetzt werden konnten. Ein lesender und schreibender Zugriff auf die Datenbasis erlaubte zudem die Integration klinischer Daten aus den Systemen der Versorgung direkt in das Dokumentationssystem. *Onkostar* kam einer stark strukturierten elektronischen Patientenakte am nächsten und eignete sich daher besonders gut zur Dokumentation der im onkologischen Bereich notwendigen hohen inhaltlichen Komplexität. Da allerdings zum Zeitpunkt der Bewertung nicht alle Funktionen getestet werden konnten und teilweise noch in der Entwicklung waren, konnte *Onkostar* nicht als elektronisches Erfassungssystem im klinischen Myelomregister verwendet werden.

Die systematische Bewertung ergab, dass die Funktionalitäten von *OpenClinica*, insbesondere die hohe Flexibilität bei Dokumentation und technischer Erweiterbarkeit, die aktive Nutzung in mehreren klinisch-wissenschaftlichen Projekten und eigene praktische Erfahrungen in Installation und Umgang, die Anforderungen des Heidelberger Myelomregisters am besten erfüllen konnte.

5.3 INTEGRATION STRUKTURIERTER INFORMATIONEN

Die Integration strukturierter Informationen aus den Systemen der Versorgung wurde für das klinische Myelomregister als ETL-Prozess für Laborparameter erprobt. Im Folgenden werden Aspekte der Planung, der Realisierung und die Ergebnisse beschrieben und bewertet.

Planung

Das EPA-System des Universitätsklinikums Heidelberg unterstützt den fallbasierten Export strukturierter gespeicherter Informationen, wie demografischer Angaben, Aufnahme-, Entlass-, Diagnose-, Prozedur- und Laboraten. Erste Analysen zeigten, dass nur wenige der exportierbaren Daten für das klinische Myelomregister relevant sind. Beispielsweise werden im klinischen Myelomregister keine OPS-Codes erfasst oder die Diagnosedaten enthielten nur den ICD-Code *C90.0 Multiples Myelom* und differenzierten nicht zwischen verschiedenen Stadien. Lediglich die Laboraten wurden in der exportierbaren Form im Merkmalskatalog der Domänen *Hämatologie* und *Immunologie* spezifiziert und für eine automatische Übernahme vorgesehen.

Die Dokumentation im klinischen Myelomregister erfolgte bisher stets aus Arztbriefen und nicht aus Originalbefunden. Zur Klärung der Fragestellung, ob sich die im EPA-System strukturiert gespeicherten und exportierbaren Laborbefunde zur Integration in das klinische Myelomre-

gister eignen, wurde eine studentische Abschlussarbeit initiiert und vom Autor dieser Dissertationsschrift betreut (Torres 2016).

Realisierung

Die Integration wurde als ETL-Prozess konzipiert und sah vor, dass exportierte Laborbefunde eingelesen, transformiert und in die Datenbank des klinischen Myelomregisters importiert wurden. Der Export musste dabei durch Eingabe der Fallnummer manuell initiiert werden. Sobald die CSV-Dateien auf ein definiertes Netzlaufwerk heruntergeladen wurden, lief der anschließende Transformationsschritt automatisch ab. Aufgrund syntaktischer und semantischer Heterogenität zwischen den CSV-Dateien und der Zielstruktur des Myelomregisters waren unterschiedliche Ansätze notwendig, um die nachfolgenden Konflikte zu lösen:

STRUKTURELLE KONFLIKTE In den exportierten Laborbefunden war nur die Fallnummer enthalten, während im klinischen Myelomregister die Zuordnung eines Befundes zum Patienten direkt über die *Patienten-Nr.* (bzw. über das *Pseudonym*) erfolgt. Entsprechend war es nicht ausreichend, nur die Laborbefunde zu exportieren, sondern die gesamten Falldaten mit *Aufnahme-* und *Entlassdatum* waren notwendig, um die Beziehung zwischen Laborbefund und Patient herzustellen. Die Falldaten wurden nur temporär zwischengespeichert und allein zur Transformation genutzt und nicht in das Myelomregister geladen.

Weiter wurde in den Laborbefunden nicht zwischen den Domänen *Hämatologie* und *Immunologie* differenziert und zusätzliche Parameter waren enthalten, die der Merkmalskatalog des Myelomregisters nicht vorsah. In einer Lookup-Tabelle wurden relevante Merkmale spezifiziert und eindeutig den beiden Domänen zugeordnet. Im Falle einer zukünftigen Erweiterung der Integration können neue Laborparameter in der Lookup-Tabelle ergänzt werden.

Zur Sicherung der Datenqualität im Myelomregister und zum Ausschluss eines Erfassungsfehlers bei der manuellen Dokumentation wird in den Domänen *Hämatologie* und *Immunologie* jeder Laborwert über ein zusätzliches Merkmal *Laborwert verfügbar* als *vorhanden* oder *fehlend* gekennzeichnet. Da dieses Merkmal in den Laborexporten nicht enthalten war, wurde es implizit gesetzt, falls ein numerischer Wert vorhanden war.

FORMATKONFLIKTE In den CSV-Dateien waren Zeitangaben als zusammenhängende Ziffernfolge definiert und wurden in das Datumsformat *yyyy-mm-dd* umgewandelt.

BENENNUNGS- UND SKALIERUNGSKONFLIKTE Unterschiedliche Bezeichnungen der Laborparameter zwischen Export und Myelomregister wurden durch eine Lookup-Tabelle aufgelöst. Zudem wurden Merkmalsdefinitionen während der Planung des klinischen Myelomregisters abgeglichen und die Laboreinheiten entsprechend dem Vorkommen in den Exporten angepasst.

Der ETL-Prozess wurde in *Talend Open Studio for Data Integration* implementiert. Die Überprüfung der Eignung der Integration erfolgte durch Vergleich mit der manuellen Transkription. Dazu wurden in einem ersten Schritt acht zufällig ausgewählte Laborparameter zu den vier Erhebungszeitpunkten *Diagnose*, *Chemotherapie*, *Transplantation* und *100 Tage nach Transplantation*, also insgesamt 32 Merkmale, von 162 zufällig ausgewählte Myelompatienten des historischen Datenbanksystems mit den im EPA-System enthaltenen Originaldokumenten abgeglichen. Dabei wurde eine manuelle Fehlerrate von 2.6 % festgestellt und die Fehler wurden für den weiteren Vergleich korrigiert.

Anschließend folgte der Export der Laborbefunde: da dabei allerdings kein bestimmter Laborbefund eines Patienten ausgewählt werden konnte, sondern immer alle Befunde zu einem Patienten exportiert werden mussten, wurden von denselben 162 Myelompatienten jeweils alle im EPA-System vorhandenen Laborbefunde exportiert. Dies führte dazu, dass beim ETL-Prozess auch Laborbefunde verarbeitet wurden, die im Myelomregister nicht erfasst oder die bei der Quelldatenverifizierung nicht überprüft wurden. Die Zuordnung der manuell transkribierten Laborbefunde mit den automatisch integrierten Befunden eines Patienten erfolgte anhand des Labordatums. Wurde bei einem Patienten dasselbe Labordatum festgestellt, unterschieden sich allerdings die Laborwerte, wurde dies als *Fehler* gewertet. Als *unvollständig* galt die automatische Integration dann, wenn ein manuell erfasster Befund nicht in der Menge der integrierten Befunde identifiziert werden konnte und auch eine Bereichssuche mit plus / minus 5 Tagen keinen Treffer brachte.

Ergebnisse

Bei Auswertung der automatisch integrierten Laborparameter wurde eine Gesamtfehlerrate, d. h. ein Unterschied zu den 32 Merkmalen der 162 Myelompatienten von 12.1 % festgestellt. Aufgeteilt in tatsächliche Fehler und unvollständige Angaben, konnten 11.4 % auf nicht vorhandene Werte bzw. auf fehlende Befunde und 0.6 % auf unterschiedliche Werte zurückgeführt werden. Während die Betrachtung der einzelnen Laborparameter keine signifikanten Unterschiede in der Fehlerverteilung ergab, zeigte sich bei den Erhebungszeitpunkten eine deutlich höhere Rate fehlender Befunde zu den Zeitpunkten *Diagnose* (29.9 %) und *Che-*

motherapie (12.7 %) gegenüber Transplantation (1.0 %) und 100 Tage nach Transplantation (2.2 %). Dies ließ sich darauf zurückführen, dass die Dokumentation im Myelomregister nicht nur intern erhobene Laborbefunde erfasste, sondern insbesondere in frühen Therapiephasen auch Befunde externer Labore oder niedergelassener Fachärzte zur Beschreibung des Krankheitszustandes heranzieht. Entsprechend sind diese Informationen nicht in digitaler und damit exportierbarer Form im EPA-System verfügbar, sondern liegen nur als eingescanntes Dokument im Archivdokumentationsystem vor. Weiter hat die Analyse der unterschiedlichen Laborwerte ergeben, dass die Originalbefunde in der Praxis nicht uneingeschränkt verwendet werden, sondern mehrere Einzelbefunde in Arztbriefen zur Interpretation eines Krankheitszustandes bewertet und zusammengefasst werden.

Bewertung und Einordnung der Ergebnisse

Technisch ist die Implementierung eines ETL-Prozesses zur Integration von Laborparametern aus dem EPA-System in das klinische Myelomregister mit überschaubarem Aufwand realisierbar. Das entwickelte Werkzeug kann für weitere Integrationsvorgänge angepasst und wiederverwendet werden. Inhaltlich weist aber die Fehlerrate der automatischen Integration von 12.1 % eine deutliche Verschlechterung gegenüber der bei der Validierung der manuellen Transkription festgestellten Fehlerrate von 2.6 % auf. Als zusätzlich negativ ist auch die Menge an überschüssigen Informationen zu nennen, die keine wissenschaftliche Relevanz für das Register hat. Weitere manuelle Zwischen- und Verifikationsschritte, die zur Selektion relevanter Informationen notwendig wären, würden allerdings wiederum den Zeitvorteil einschränken und nicht das Problem extern erhobener und nur in digitalisierter Form vorliegender Informationen lösen. Beide Gründe führten zu der Schlussfolgerung, dass sich die Integration strukturierter Informationen aus dem EPA-System in das klinische Myelomregister aktuell nur eingeschränkt eignet. In klinischen Registern, die ausschließlich interne Informationen der eigenen Versorgungseinrichtung nutzen, stellt der ETL-Prozess hingegen eine valide Möglichkeit der Informationsintegration dar.

5.4 INTEGRATION UNSTRUKTURIERTER INFORMATIONEN

Die Integration unstrukturierter Informationen aus den Systemen der Versorgung wurde für das klinische Myelomregister für hämatologisch-onkologische Arztbriefen als Text Mining erprobt. Aspekte, die bei Planung und Realisierung relevant waren, werden im Folgenden beschrieben und die Ergebnisse dargestellt und bewertet.

Planung

Im klinischen Myelomregister werden Informationen zur Routinebehandlung überwiegend aus Arztbriefen erfasst, die den Krankheits- und Therapieverlauf eines Myelompatienten detailliert und umfassend darstellen. Da sie fachärztlich erstellt, mehrfach gelesen und geprüft werden und das Behandlungsgeschehen aus medizinischer Sicht wiedergeben und verifizieren, bilden sie eine geeignete Dokumentationsgrundlage. Am Universitätsklinikum Heidelberg werden Arztbriefe elektronisch im EPA-System angelegt und sind in der Sektion Multiples Myelom unter anderem in die Abschnitte *Hauptdiagnose*, *Nebendiagnosen*, *Prognosefaktoren*, *Therapie*, *Verlauf* und *Epikrise* gegliedert. Innerhalb der einzelnen Abschnitte werden Freitexte verwendet, die eine repräsentative Abbildung und Dokumentation komplexer Sachverhalte erlauben. Der Zugriff auf das EPA-System ist nur über Benutzerschnittstellen möglich, und entsprechend sind manuelle Transkriptionsvorgänge durch medizinische Dokumentare eine wichtige Möglichkeit klinische Informationen aus Arztbriefen in das Myelomregister zu übertragen. Dies ist allerdings ein fehleranfälliger sowie zeit- und ressourcenaufwändiger Prozess. Methoden des Information Retrieval (IR), die mit Hilfe algorithmischer Klassifikationsverfahren Informationen in Textpassagen erkennen und extrahieren, wurden als eine Möglichkeit zur Unterstützung der manuellen Transkription angesehen. Inwieweit sich diese Verfahren eignen, Merkmale des Myelomregisters automatisch aus hämatologisch-onkologischen Arztbriefen zu erfassen, wurde gemeinsam mit dem *Institut für Computerlinguistik der Universität Heidelberg* in einer durch den Autor dieser Dissertationsschrift betreuten studentischen Abschlussarbeit systematisch untersucht (Krauss 2015). Das Ziel war dabei nicht die Realisierung eines einzelnen Klassifikationsverfahrens, sondern der Vergleich mehrerer etablierter Verfahren und Methoden hinsichtlich ihrer Eignung und Güte.

Realisierung

Der Vergleich mehrerer Verfahren des IR sah die Entwicklung eines Frameworks vor, mit dem unterschiedliche Vorverarbeitungsschritte und Klassifikatoren kombiniert werden konnten. Als Klassifikatoren wurden ein *Maximum-Entropie Klassifikator* und eine *Support Vector Machine* unter Verwendung der frei verfügbaren Java Bibliotheken *MALLET* und *LIBSVM* umgesetzt (Chang u. Lin 2011; McCallum 2002). Methoden der Vorverarbeitung, die die Klassifikationsergebnisse durch automatische Textaufbereitung verbessern sollten, waren ein eigenes Lexikon zur Auflösung medizinischer Abkürzungen, mehrere syntaktische Regeln zur Erkennung regulärer Ausdrücke, eine Rechtschreibkorrektur durch die Wörterbücher *Hunspell* und *UMLS* sowie das Erkennen und Hervorhe-

ben von Nominalphrasen in Textpassagen (NP-Chunking) über die Java Bibliothek *OpenNLP* (Lindberg et al. 1993; Morton et al. 2005; Németh 2014). Die beiden Klassifikatoren wurden ohne Vorverarbeitungsschritt sowie mit jeweils einer der vier vorgeschalteten Methoden untersucht, was in insgesamt zehn verschiedenen Kombinationen resultierte.

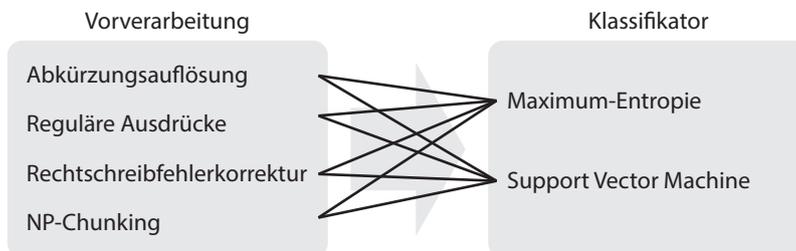


Abbildung 18: Kombinationen mehrerer Methoden der Vorverarbeitung mit den zwei Klassifikatoren zur Prüfung der Eignung und Güte eines automatischen IR zur Extraktion unstrukturierter Informationen aus hämatologisch-onkologischen Arztbriefen.

Zum Training und Testen der Klassifikatoren wurde bei 737 zufällig ausgewählten hämatologisch-onkologischen Arztbriefen der Sektion Multiples Myelom der Textabschnitt *Hauptdiagnose* extrahiert und die relevanten Merkmale wurden manuell transkribiert. Da in einem Textabschnitt mehrere Einzeldiagnosen enthalten waren, ließen sich 867 Diagnosen mit jeweils bis zu sechs ableitbaren Merkmalen unterscheiden. Diese codierten unter anderem die *Diagnose*, den *Antikörper-* und *Leichtkettentyp* sowie das *Krankheitsstadium nach Durie-Salmon*. Um Fehler bei der manuellen Transkription auszuschließen, erfolgte diese durch zwei unabhängige Personen. 90 % der Diagnosen wurden zum Training der Klassifikatoren verwendet und die restlichen 10 % zur automatischen Klassifikation. Zum Vergleich automatischer und manueller Transkription wurde pro Merkmal *Fehlerrate* und *F-Maß* als harmonisches Mittel aus Genauigkeit und Trefferquote berechnet. Die Aufteilung in Trainings- und Testdiagnosen erfolgte zufällig, wurde 100-fach wiederholt und die Durchschnittswerte der jeweils bestimmten Maßzahlen wurden bestimmt.

Ergebnisse

Die untersuchten Klassifikatoren wiesen bereits ohne Vorverarbeitungsschritt eine geringe Fehlerrate auf, die in Abhängigkeit des betrachteten Merkmals zwischen 0,95 % für die *Diagnose* und 4,16 % für den *Leichtkettentyp* lag. Die Fehlerrate der Support Vector Machine war bei vier von sechs Merkmalen geringer im Vergleich zum Maximum-Entropie Klassifikator. Die Hinzunahme der Methoden der Vorverarbeitung wirk-

te sich sowohl positiv, als auch negativ auf die Ergebnisse der einzelnen Merkmale aus und ließ keinen eindeutigen Trend erkennen. Beispielsweise verringerte sich die Fehlerrate des *Krankheitsstadiums* durch Auflösung medizinischer Abkürzungen signifikant beim Maximum-Entropie Klassifikator, führte allerdings bei der Support Vector Machine zu einer Erhöhung. Auch verhielten sich nicht alle Merkmale bei einer Kombination aus Klassifikator und Vorverarbeitungsschritt gleich. So verbesserte die Rechtschreibkorrektur bei der Support Vector Machine die Merkmale *Antikörper-* und *Leichtkettentyp* signifikant, verschlechterte hingegen die *Diagnose*. Das F-Maß war von ähnlichen Schwankungen betroffen, lag aber insgesamt bei der Support Vector Machine höher.

Bewertung und Einordnung der Ergebnisse

Die Ergebnisse der Untersuchung zeigten, dass sich IR zur Extraktion von Diagnosemerkmalen aus hämatologisch-onkologischen Arztbriefen grundsätzlich eignet und durch merkmalspezifische Anpassungen weitere Optimierungen möglich sind. Das entwickelte Framework wurde generisch erstellt und ist auf andere Krankheitsentitäten übertragbar, erfordert allerdings geeignete Trainingsdatensätze. Problematisch hingegen ist, dass ein Verfahren nicht eindeutig einem anderen vorgezogen werden kann und sich unterschiedlich auf die jeweils zu klassierenden Merkmale auswirkt. Zudem bestehen Arztbriefe aus mehreren Abschnitten und erfordern zur vollständigen Analyse und Erfassung die Auswahl eines geeigneten Verfahrens pro Abschnitt.

Erlauben die Systeme der Versorgung keinen elektronischen Zugriff auf Quelldokumente, sind manuelle Zwischenschritte nötig, bei dem relevante Textpassagen zunächst identifiziert und manuell in ein Text Mining Werkzeug eingefügt werden müssen. Für eine praktische Umsetzung wäre es zudem sinnvoll, die automatisch erzeugten Ergebnisse vor der Übernahme in das Register dem Benutzer anzuzeigen, sodass dieser sie im Fehlerfall korrigieren kann. Da bei diesem Vorgehen der Zeit- und Ressourcenaufwand als kaum geringer zu einer vollständig manuellen Erfassung durch einen erfahrenen medizinischen Dokumentar eingeschätzt wird, wurde das IR im Myelomregister nicht in die Routinedokumentation übernommen. Für klinische Register, die jedoch elektronisch auf digitale Dokumente und deren Inhalt zugreifen können, sind die untersuchten Verfahren als unterstützende Maßnahme der manuellen Transkription zu werten.

5.5 METADATENMANAGEMENTSYSTEM KLINISCHES MYELOMREGISTER

Zur Unterstützung der Integration wissenschaftlicher Datensammlungen in das klinische Myelomregister wurde ein Metadatenmanagementsystem (MDMS) entwickelt und die Metadaten der im Registerprotokoll identifizierten Systeme und Merkmale wurden eingepflegt. Aspekte zur Planung und Umsetzung eines MDMS und die Bewertung der Ergebnisse werden im Folgenden beschrieben.

Planung

Die Integration historischer Datensammlungen der Sektion Multiples Myelom in das klinische Myelomregister erforderte die Analyse der vorhandenen Merkmale, ihrer Definitionen und Ausprägungen. Dazu wurden Syntax und Semantik der Merkmale untersucht sowie konzeptionell ähnliche Merkmale aufeinander abgebildet. Die Implementierung eines MDMS gemäß Standard *ISO / IEC 11179* erfolgte in einer durch den Autor dieser Dissertationsschrift betreuten studentischen Abschlussarbeit (Jones 2014). Die konzeptionelle Erfassung und Einordnung der Merkmale fand im Rahmen eines Praktikums unter Einbeziehung schriftlicher Dokumentationsanleitungen und ergänzender mündlicher Erläuterungen statt.

Realisierung

Für die Umsetzung des MDMS wurde das in Kapitel 4.2.6 hergeleitete physische Datenmodell implementiert. Als Metadaten wurden die, in dem Registerprotokoll aufgeführten Eigenschaften von Merkmalen verwendet. Diese wurden um weitere Metadaten ergänzt, wie die *Datensammlung* und der *Erhebungszeitraum*. Das Datenmodell wurde mit dem relationalen DBS *PostgreSQL* und die Benutzeroberfläche mit dem *Google Web Toolkit* umgesetzt. Der Zugriff auf die Weboberfläche des MDMS wurde nur für authentifizierte Benutzer innerhalb des Universitätsklinikums Heidelberg vorgesehen.

Ergebnisse

Das relationale Datenmodell des MDMS besteht aus vier Tabellen und bildet die Beziehungen zwischen *Konzept*, *Merkmal*, *Wertebereich* und *Wert* ab. Abbildung 19 zeigt das Datenmodell, die spezifizierten Eigenschaften und jeweiligen Datentypen. Instanzen jeder Tabelle werden eindeutig über eine numerische *Identifikationsnummer* (ID) definiert. Die weiteren

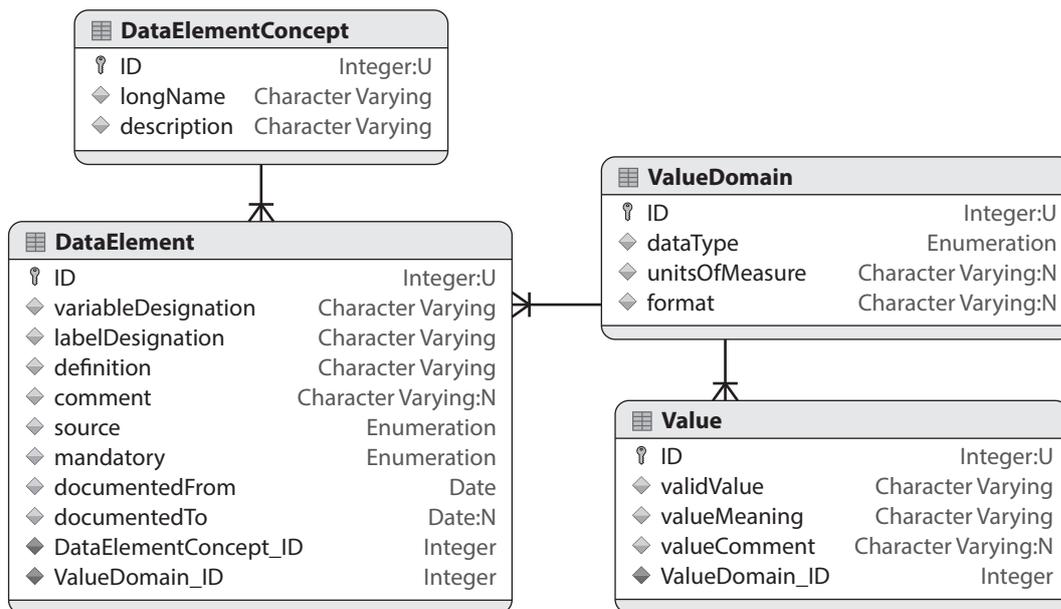


Abbildung 19: Relationales Datenmodell des MDMS für das klinische Myelomregister.

Eigenschaften dienen dazu, ein Merkmal eindeutig und umfassend zu beschreiben und einem Konzept zuzuordnen. Für die Tabelle Konzept (*DataElementConcept*) wurden die Eigenschaften ausführliche Bezeichnung (*longName*) und Beschreibung (*description*) festgelegt, die für alle zuzuordnenden Merkmale identisch sein müssen. Unterschieden sich zwei Merkmale in ihrer Beschreibung, dann tun sie dies auch in ihrem Konzept. Jedes Merkmal muss genau einem Konzept zugeordnet werden, dessen Bezeichnung im einfachsten Fall identisch mit der Bezeichnung des Merkmals auf der Oberfläche (*labelDesignation*) ist. Ein Merkmal (*DataElement*) lässt sich eindeutig charakterisieren durch einen Variablennamen (*variableDesignation*), eine natürlichsprachige Bezeichnung auf der Benutzeroberfläche (*labelDesignation*), eine Definition (*definition*), einen Kommentar (*comment*), eine Quellangabe (*source*), die Angabe ob es sich um ein Pflichtfeld handelt (*mandatory*) und einen Erhebungszeitraum (*documentedFrom* und *documentedTo*). Jedem Merkmal ist ein Wertebereich (*ValueDomain*) zugeordnet, der den Datentyp (*dataType*), die Maßeinheit (*unitsOfMeasure*) und das Format (*format*) spezifiziert. Setzt sich ein Wertebereich aus mehreren Werten zusammen, wie bei einer Aufzählung, dann werden die einzelnen Optionen als Wert (*Value*) definiert. Ein Wert besteht dabei aus seiner Codierung (*validValue*), der Bedeutung (*valueMeaning*) und einem optionalen Kommentar (*valueComment*).

Die Erfassung der Merkmale aus den historischen Datensammlungen in das MDMS erfolgte nach Analyse der Ausprägungen, der Dokumen-

tationsanleitungen und Rücksprachen mit den Systemanwendern. Gemeinsame Konzepte und die Verknüpfung mehrerer Merkmale zu einem Konzept wurden in einem zweiten Schritt erstellt. Insgesamt konnten ca. 660 Merkmale erfasst und auf 450 Konzepte abgebildet werden. Den Merkmalen wurden 300 Wertebereiche mit ungefähr 600 einzelnen Werten zugeordnet. Dadurch ließ sich unter anderem die Änderung der Definition eines Merkmals (bzw. eines Konzepts) im Zeitverlauf nachvollziehen. Beispielsweise wurde das Konzept *Diagnose* in der historischen Excel-Datei als dichotomes Merkmal *Multiple Myelom ja / nein* erhoben und später hinsichtlich der Optionen *MGUS*, *Multiple Myelom* und *Plasmazell-Leukämie* unterschieden.

Bewertung und Einordnung der Ergebnisse

Die praktische Umsetzung des MDMS nach dem Standard *ISO / IEC 11179* zeigte die Machbarkeit einer Implementierung als relationales Datenbanksystem (DBS). Das System kann als eigenständige Anwendung auf für andere ausgeführt werden und ist ohne Anpassungen nutzbar. Problematisch bei der Implementierung war die Realisierung einer einheitlichen und intuitiven Benutzeroberfläche über die Metadaten und Zuordnungen geeignet dargestellt und miteinander in Beziehung gesetzt werden konnten. Die systematische Erfassung und Analyse der historisch erhobenen Merkmale war während der Planungsphase des klinischen Myelomregisters hilfreich und lieferte wertvolle Überblicke in die vorhandenen Datensammlungen, deren Eigenschaften und Inhalte. Da die Anzahl tatsächlicher Merkmale aufgrund von Hilfs- und Informationsvariablen nicht immer mit der vorhandenen Spalten- und Attributanzahl übereinstimmte, konnte über das MDMS die Menge eindeutiger Konzepte ermittelt werden. Auch war es möglich, Metadaten über Queries so abzufragen, dass die einem Konzept zugeordneten Merkmale und Wertebereiche ausgegeben und als Vorstufe der Integration hinsichtlich ihres Harmonisierungsverhaltens geprüft werden konnten. Vorgesehen ist, das MDMS stärker in den ETL-Integrationsprozess mit molekularbiologischen Daten der Zytogenetik-Forschungsdatenquelle einzubinden und die Funktionalität, beispielsweise um die Ausgabe von Merkmalskatalogen, zu erweitern.

5.6 ABSCHLIESSENDE BEMERKUNGEN

Die informationsverarbeitende und speichernde Infrastruktur des klinischen Myelomregisters wurde gemäß der Referenz-IT-Architektur umgesetzt. Alle Anwendungssysteme sind im Registerprotokoll (Protokollpunkt 6.3) spezifiziert.

Als prolektives Dokumentationssystem wurde mit *OpenClinica* ein modernes elektronisches Datenerfassungssystem ausgewählt, dessen primäres Einsatzgebiet bei klinischen Studien liegt. Aufgrund der Abbildung individueller Krankheits-, Behandlungs- und Therapieverläufe ist es allerdings auch für klinische Register geeignet. *OpenClinica* bietet zur Dateneingabe eine webbasierte Benutzerschnittstelle sowie weitere Schnittstellen zur Integration. Da die Patientendokumentation nur pseudonymisiert möglich ist, wurde mit *REDCap* ein weiteres elektronisches Erfassungssystem zur Verwaltung identifizierender Daten genutzt. Die gemeinsame Ansicht klinischer und identifizierender Daten erfolgte über den Portalserver *Liferay* in Form einer elektronischen Patientenliste.

Bezüglich der Integration lag der Schwerpunkt beim Heidelberger Myelomregister auf der Unterstützung manueller Transkriptionsvorgänge durch automatisierte Verfahren zur Übernahme strukturierter und unstrukturierter Informationen aus dem KIS. Zusätzlich waren retrolektive Daten in historischen Datensammlungen vorhanden, die vor der Integration mit Hilfe eines MDMS analysiert und zueinander in Beziehung gesetzt werden konnten, sodass sich daraus Harmonisierungs- und Transformationsregeln ableiten ließen. Retrolektive Daten wurden vor der Migration in das klinische Register geprüft und mit den Originaldaten im KIS abgeglichen. Die Integration wurde der Referenz-IT-Architektur entsprechend in einem klinischen Data Warehouse umgesetzt und neben klinischen auch molekularbiologische Daten ergänzt.

DISKUSSION

In dieser Dissertationsschrift wurde ein Vorgehensmodell für die systematische Planung eines klinischen Registers entwickelt. Dieses kann als Registerprotokoll zur einheitlichen, angemessenen und nachvollziehbaren Herleitung von Zielen, Aufgaben und Fragestellungen eines klinischen Registers dienen, sodass daraus systematisch die Anforderungen an die Realisierung des Registersystems abgeleitet werden können. Aufbauend auf dem Vorgehensmodell wurde eine Referenz-IT-Architektur entwickelt, die sich für die Implementierung und den Betrieb eines flexiblen und erweiterbaren klinischen Registers eignet. Vorgehensmodell und Referenz-IT-Architektur wurden unter besonderer Berücksichtigung der Integration der Anwendungssysteme des klinischen Registers in die vorhandene Informationsinfrastruktur einer klinisch-wissenschaftlichen Versorgungseinrichtung entwickelt.

Mit dem Vorgehensmodell und der Referenz-IT-Architektur liegt ein umfassender allgemeiner Leitfaden für alle Entwicklungsphasen eines klinischen Registers vor. Der Leitfaden geht über existierende Empfehlungen für klinische Register hinaus und betrachtet zusätzlich systematisch verschiedene Integrationsaspekte.

Die Referenz-IT-Architektur setzt sich aus fünf Komponenten zusammen: einem prolektiven Dokumentationssystem, den vorhandenen Systemen von Versorgung und Forschung, einem Data Warehouse sowie einem Metadatenmanagementsystem. Für diese Komponenten wird im Leitfaden bewusst keine spezifische Softwarearchitektur vorgeschlagen, da abhängig von dem Krankheitsbild und den Rahmenbedingungen des zu implementierenden Registers unterschiedliche Lösungsansätze denkbar sind. Ziel des Leitfadens ist es, dediziert diese Alternativen aufzuzeigen und Hinweise zu geben, welche Kriterien als Entscheidungsgrundlage genutzt werden können. Sind beispielsweise mehrere Datenquellen für die Nutzung im klinischen Register vorgesehen, so kann die Implementierung eines Metadatenmanagementsystems gemäß dem Standard *ISO / IEC 11179* sinnvoll sein, da Konzepte und Wertebereiche heterogener Datensammlungen aufeinander abzubilden sind. Für die Implementierung eines elektronischen Datenerfassungssystems können sowohl kommerzielle als auch frei verfügbare Systeme oder Eigenentwicklungen sinnvoll sein. Diese Entscheidung hängt zum Beispiel wesentlich davon ab, ob ein dynamisches Dokumentationsmodell für eine flexible Erfassung individueller Krankheitsverläufe notwendig ist oder nicht.

Damit der Leitfaden direkt als Planungsgrundlage für neue klinische Register genutzt werden kann, wurde Kapitel 4 dieser Dissertationsschrift so ausformuliert, dass es auch außerhalb dieser Arbeit ein eigenständiges und abgeschlossenes Dokument darstellt. Über diesen Leitfaden hinaus sind für einen reibungslosen Betrieb auch organisatorische Anforderungen an kollaboratives Arbeiten oder zu Trägerschaft, Finanzierung, ethischen oder rechtliche Voraussetzungen von großer Bedeutung. Hierauf gehen die Arbeiten von Gliklich und Dreyer (Gliklich u. Dreyer 2014) und Müller et al. (Müller et al. 2010) und die dort referenzierte Literatur weiter ein.

Der in dieser Dissertationsschrift entwickelte Leitfaden wurde für die Planung und den Aufbau des neu am Universitätsklinikum Heidelberg einzuführenden klinischen Myelomregisters angewendet. Bei der Realisierung des Registers entstanden folgende weitere Ergebnisse, die zusätzlich zum Leitfaden auch für den Aufbau anderer Register genutzt werden können:

- Ein Kriterienkatalog zur Beurteilung der Eignung von Softwareprodukten zur elektronischen Datenerfassung für ein prolektives klinisches Register
- Ein Werkzeug zur Integration strukturierter Informationen aus dem rechnerunterstützten Teil eines Krankenhausinformationssystems in ein prolektives klinisches Register
- Ein Werkzeug zur Integration unstrukturierter Informationen aus hämatologisch-onkologischen Arztbriefen in ein prolektives klinisches Register
- Ein Metadatenmanagementsystem zur Beschreibung von Merkmalen und zur Abbildung semantischer und syntaktischer Beziehungen zwischen den Merkmalen heterogener Systeme.

Im Folgenden werden das Vorgehensmodell zur systematischen Planung und die Referenz-IT-Architektur im Detail diskutiert.

6.1 VORGEHENSMODELL ZUR SYSTEMATISCHEN PLANUNG

Zur Entwicklung des Vorgehensmodells wurden methodische Richtlinien und Handlungsanweisungen für Planung und Aufbau klinischer Register systematisch recherchiert und bewertet. Dabei wurde festgestellt, dass die vorhandenen Veröffentlichungen unterschiedliche Aspekte der Registerentwicklung behandeln und kein umfassendes systematisches Vorgehen beschreiben. Die Literaturrecherche wurde auf Leitfäden zu Registern mit wissenschaftlichem Anwendungsgebiet eingeschränkt, die in

Deutschland genutzt werden. Register, die gesundheitsökonomische oder sozialpolitische Zielsetzungen verfolgen, konnten dadurch ausgeschlossen werden. Durch die Einschränkung auf eine Verwendung in Deutschland sollte erreicht werden, dass die Leitfäden in der Lage sind, nationale Regelungen zu berücksichtigen, die die Entwicklung klinischer Register beeinflussen wie beispielsweise die Datenschutzgesetzgebung. Um hier einer Verzerrung durch eine zu starke Einschränkung entgegenzuwirken, wurde die Recherche um das Schneeballsystem erweitert. Dadurch wurden mit den Veröffentlichungen von Gliklich und Dreyer und Mathis und Wild auch ausländische, aber explizit für klinisch-wissenschaftliche Register geeignete Leitfäden gefunden (Gliklich u. Dreyer 2014; Mathis u. Wild 2008).

Das Vorgehensmodell wurde an das Dokumentationsprotokoll von Leiner und Haux angelehnt und um register- sowie integrationsspezifische Aspekte erweitert (Leiner u. Haux 1996). Das damit vorgeschlagene Registerprotokoll sieht vor, dass Registerziele aus konkret zu lösenden Problemen abgeleitet werden und anschließend geprüft wird, ob Daten aus vorhandenen Systemen eine prolektive Erhebung sinnvoll ergänzen können. Entsprechend wird die Integration in vorhandene Systeme von Versorgung und Forschung frühzeitig in der Planungsphase berücksichtigt und an verschiedenen Stellen im Registerprotokoll aufgegriffen. Das erarbeitete Vorgehensmodell ist aber auch für die Planung klinischer Register mit rein prolektiver Datenerhebung geeignet und betrachtet gegenüber dem Dokumentationsmodell von Leiner und Haux zusätzliche registerspezifische Anforderungen wie die Organisation des Follow-Up und Maßnahmen zur Sicherung der Datenqualität. Die Nutzung des Vorgehensmodells für reine Integrationslösungen ohne prolektive Datenerhebung ist ebenfalls möglich, sofern diese dadurch motiviert sind, dass die Integration problemorientiert und zielgerichtet zur Beantwortung bestehender wissenschaftlicher Fragestellungen erfolgt.

Aufgrund der großen Bedeutung von klinischen Registern für die medizinische Forschung sind in den letzten Jahren einige Publikationen entstanden, die für spezifische Aspekte des Aufbaus und Betriebs von Registern wichtig sind wie beispielsweise Qualitätsindikatoren oder informationstechnische Anforderungskataloge (Michalik et al. 2014; Nonnemacher et al. 2007). Diese stellen wertvolle Ergänzungen zum Registerprotokoll dar und bieten eine weitergehende Unterstützung bei der Detailplanung einzelner durch das Registerprotokoll vorgeschlagener Schritte.

Die Eignung des Vorgehensmodells konnte durch die Erstellung des konkreten Registerprotokolls für die Planung des klinischen Myelomregisters des Universitätsklinikums Heidelberg und dessen anschließender Realisierung gezeigt werden. Es ermöglichte frühzeitig eine Diskussionsgrundlage für alle Beteiligten und die schriftliche Fixierung wichtiger

Entscheidungen. Dabei war zu beobachten, dass es für die unterschiedlichen beteiligten Interessensgruppen durchaus herausfordernd war, sich zunächst auf relativ abstraktem Niveau auf Probleme und Zielsetzungen des Registers zu verständigen und erst später im Planungsprozess konkrete wissenschaftliche Fragen und Merkmale abzuleiten. Daher empfehlen sich eine enge Begleitung des Prozesses, der Erstellung und Pflege des Registerprotokolls sowie der Vor- und Nachbereitung von Diskussionsrunden durch eine Person, die nicht unmittelbar eigene wissenschaftliche Fragestellungen mit der Auswertung des Registers verfolgt.

6.2 REFERENZ-IT-ARCHITEKTUR

Ausgehend von den Anforderungen an klinische Register und dem erarbeiteten Vorgehensmodell wurde eine Referenz-IT-Architektur entwickelt, die nicht nur das prolektive Dokumentationssystem umfasst, sondern insbesondere die Integration in vorhandene Systeme von Versorgung und Forschung ermöglicht. Dabei wurden mehrere Anwendungssysteme und Kommunikationskomponenten systematisch eingeführt und Entscheidungen aufgezeigt, die bei Auswahl und Entwicklung der informationstechnischen Infrastruktur eines klinischen Registers getroffen werden sollten. Die Referenz-IT-Architektur stellt ein abstrahiertes Modell der logischen Werkzeugebene dar, das unterschiedliche Implementierungsvarianten zulässt. Ausgangspunkt ist ein prolektives Dokumentationssystem als zentrales Anwendungssystem, das mit verschiedenen Integrationskomponenten interagiert. Diese Integrationskomponenten sind optional und erlauben somit den Aufbau verschiedener klinischer Register mit unterschiedlichen Stufen an Integrationsanforderungen.

In der Literatur wird aktuell die Eignung von Studiendatenmanagementsystemen für klinische Register geprüft (Lindörfer u. Mansmann 2015; Michalik et al. 2014). Im Verlaufe dieser Dissertationsschrift hat sich gezeigt, dass klinische Register eigene Anforderungen stellen wie beispielsweise eine hohe inhaltliche und technische Flexibilität hinsichtlich der Erfassung individueller Krankheitsverläufe sowie der Adaption auf neue Rahmenbedingungen, Zielsetzungen und Fragestellungen. Daher wurden Bewertungskriterien zur Beurteilung elektronischer Erfassungssysteme, die die Anforderungen klinischer Register zusätzlich berücksichtigen, hergeleitet und als Kriterienkatalog zur Verfügung gestellt. Aus diesem Grund wird in der Arbeit ein prolektives Dokumentationssystem für klinische Register auch explizit als *Registerdatenmanagementsystem* bezeichnet und gegenüber Studiendatenmanagementsystemen abgegrenzt.

Die Integration des klinischen Registers in die vorhandenen Systeme der Versorgung wurde in der Referenz-IT-Architektur als Transkription klinischer Informationen in ein prolektives Dokumentationssystem vorgesehen. Alternative IT-Infrastrukturen für klinische Register wie von

Kasper et al. und Ganslandt et al. publiziert, sehen hingegen direkt die Integration in ein Data Warehouse vor und benötigen kein dediziertes Datenerfassungssystem (Ganslandt et al. 2011; Kaspar et al. 2016). Diese Ansätze sind insofern attraktiv, als dass dieselben klinischen Daten unmittelbar sowohl für Patientenversorgung als auch für klinische Forschung genutzt werden können, erfordern allerdings eine strukturierte und merkmalsbasierte Erhebung und Erfassung direkt in den Systemen der Versorgung. Liegen nicht alle Daten strukturiert und elektronisch in einem Krankenhausinformationssystem vor, ist auch die direkte Integration in ein Data Warehouse nicht möglich und eine (manuelle) Transkription wird notwendig. In der Referenz-IT-Architektur wurde das elektronische Dokumentationssystem als *Verbindung* zwischen Krankenhausinformationssystem und klinischem Data Warehouse spezifiziert, sodass auch externe, papierbasierte und unstrukturierte Informationen systematisch erfasst und verarbeitet werden können. Das Data Warehouse wird als zentrale Plattform zur Integration prolektiver Registerdaten mit den retrolektiven Systemen der Forschung eingesetzt, da hierbei von einem höheren Maß an Strukturierung und Standardisierung gegenüber den Systemen der Versorgung ausgegangen werden kann.

Der Integrationsprozess kann von einem Metadatenmanagementsystem zur Abbildung semantischer Beziehungen heterogener Systeme profitieren. Durch die Umsetzung nah am Standard *ISO / IEC 11179* wird zwar die Komplexität zur Beschreibung von Merkmalen erhöht, allerdings geht damit auch eine höhere Flexibilität bei Konzeptualisierung und Spezifikation der Merkmale einher. Diese erwies sich zur Repräsentation semantischer Ähnlichkeiten und Unterschiede zwischen den Merkmalen vorhandener Datensammlungen als zielführend und erlaubte Harmonisierungs- und Transformationsregeln direkt herzuleiten. Integrationsbestrebungen wie von Firnkorn et al. beschrieben, die tabellarische Dokumente zur Beschreibung heterogener Systeme und Datensammlungen verwenden, können durch systematisch erfasste Metadaten gemäß dem dargelegten physischen Datenmodell sinnvoll ergänzt werden (Firnkorn et al. 2015).

Die Referenz-IT-Architektur war die Grundlage für die Entwicklung der Anwendungssysteme des klinischen Myelomregisters des Universitätsklinikums Heidelberg. Herausforderungen an das prolektive Dokumentationssystem waren insbesondere die Kombination aus standardisierter Erfassung und individueller Abbildung der Versorgungsrealität. Keines der geprüften Erfassungssysteme konnte alle Anforderungen umfänglich erfüllen. In diesem Fall kann eine Eigen- oder Auftragsentwicklung erwogen werden, wie in den Veröffentlichungen von Min et al. und Plotnicki et al. beschrieben (Min et al. 2011; Plotnicki et al. 2013). Im Falle des Heidelberger Myelomregisters hat man sich wegen der hohen zu erwartenden Aufwände bezüglich Entwicklung, Wartung und Pflege jedoch

gegen eine Eigenentwicklung und für einen Kompromiss zwischen Anforderungen und möglicher Funktionalität entschieden. Die Integration des klinischen Myelomregisters wurde insbesondere zur Unterstützung der manuellen Transkription mit den Systemen der Versorgung vorgesehen. Einschränkungen beim Zugriff auf den rechnerbasierten Teil des Krankenhausinformationssystems trugen allerdings dazu bei, dass jeweils manuelle Zwischenschritte nötig waren und vollautomatisierte Lösungen nicht umgesetzt werden konnten. Bei Systemen der Versorgung, die wie bei Kasper et al. und Ganslandt et al. eine strukturiertere Erfassung und einen elektronischen Zugriff auf klinische Daten erlauben (Ganslandt et al. 2011; Kaspar et al. 2016), können mit den entwickelten Verfahren zur Integration unstrukturierter Informationen aus Arztbriefen weitere Merkmale automatisch extrahiert und in ein klinisches Data Warehouse integriert werden.

6.3 AUSBLICK

In wissenschaftlich ambitionierten Versorgungseinrichtungen ist es heute üblich, dass klinische Informationen sowohl als Versorgungsdaten im Krankenhausinformationssystem und auch als Forschungsdaten in einer Forschungsdatenbank gespeichert werden. Diese Daten werden meist doppelt erfasst, obwohl bereits vor fünf Jahrzehnten eine multiple Verwendung als sinnvoll erachtet wurde (Yoder et al. 1964). Aktuelle Ansätze zur multiplen Verwendung von Daten aus der Versorgung für die Forschung unterstützen beispielsweise die Rekrutierung von Patienten für klinische Studien oder ergänzen und verbessern deren Ergebnisse (Hemkens et al. 2016; Kreuzthaler et al. 2015). Weitergehende Ansätze erweisen sich auch heute noch in der Praxis meist als schwer umsetzbar. Dies liegt wesentlich daran, dass verfügbare Systeme bei der Abbildung versorgungs- und realitätsnahe Erfassungen bereits jetzt an Grenzen stoßen und somit den Bedarf nach stark strukturierten elektronischen Patientenakten wecken (Drepper u. Semler 2016). Von daher ist zu beobachten, ob sich in Zukunft Systeme zur Unterstützung der Patientenversorgung und der klinischen Forschung annähern werden und den Anforderungen beider Seiten gleichzeitig gerecht werden können. Die Erfahrungen in dieser Arbeit haben gezeigt, dass dafür vor allem weitergehende Erkenntnisse auf dem Gebiet der Repräsentation und Verarbeitung longitudinaler Daten notwendig sind (Capurro et al. 2015). Bisher ist zum Beispiel bei der Übernahme von Labordaten meist eine manuelle Auswahl der in dem wissenschaftlichen Kontext gewünschten Untersuchung notwendig. Auch kann weitere Forschung über die Nutzung von Text Mining für die Extraktion von Informationen aus natürlichsprachigen Texten dazu beitragen, dass Daten aus der Patientenversorgung in klinischen Registern genutzt werden können.

Die klinische Forschung ist heute noch stark durch ein problemorientiertes Vorgehen und hypothesengetriebene Studien geprägt. Die Auswertung großer Datenmengen gewinnt allerdings in letzter Zeit stark an Bedeutung und es stellt sich die Frage, ob in Zukunft auch mit datengetriebenen Methoden Evidenz für die Medizin generiert werden kann (Binder u. Blettner 2015; Rüping u. Graf 2013). Daher wird es beim Aufbau klinischer Informationssysteme in Zukunft verstärkt notwendig sein, einen problem- und einen datenorientierten Ansatz gemeinsam zu betrachten. Mit dem entwickelten Leitfaden wurde ein erster Schritt in diese Richtung gegangen und datenorientierte Ansätze werden in dem Sinne berücksichtigt, dass der Integration existierender Datensammlungen eine wichtige Rolle beim Aufbau klinischer Register zudedacht wurde.

ZUSAMMENFASSUNG

Klinische Register können die Forschungs- und Versorgungslücke zwischen evidenzbasierten klinischen Studien einerseits und der Alltagssituation andererseits schließen, indem Patienten aus der Versorgungsrealität eingeschlossen und therapeutische Langzeiteffekte, Nebenendpunkte und Begleiterscheinungen beobachtet werden. Eine systematische Planung und ein standardisiertes Vorgehen gelten dabei als wichtige Faktoren für den unmittelbaren und langfristigen Erfolg eines klinischen Registers. Methodische Richtlinien und Handlungsanleitungen hierfür existieren allerdings kaum und vorhandene Konzepte basieren auf der Annahme, dass mit einem neuen klinischen Register auch neue wissenschaftliche Dokumentationen und Infrastrukturen aufzubauen und einzuführen sind. Entsprechend bleiben bestehende Datenbestände und Informationssysteme bislang meist unberücksichtigt.

Ausgehend von einer Literaturrecherche und den daraus abgeleiteten Anforderungen an ein klinisches Register und dessen informationstechnische Infrastruktur wurden ein Vorgehensmodell zur systematischen Planung und eine Referenz-IT-Architektur zum Aufbau und Betrieb einer flexiblen und erweiterbaren IT-Infrastruktur entwickelt. Das Vorgehensmodell kann als Registerprotokoll zur einheitlichen, angemessenen und nachvollziehbaren Beschreibung und zur Herleitung von Zielen, Aufgaben und Fragestellungen eines klinischen Registers eingesetzt werden. Die Referenz-IT-Architektur besteht aus fünf Komponenten: einem prolektiven Dokumentationssystem, den vorhandenen Systemen von Versorgung und Forschung, einem Data Warehouse sowie einem Metadatenmanagementsystem. Es werden Entscheidungen hinsichtlich des Zusammenwirkens spezifiziert, die bei Auswahl und Entwicklung getroffen werden sollten, sodass unterschiedliche Implementierungsvarianten möglich sind. Vorgehensmodell und Referenz-IT-Architektur wurden unter besonderer Berücksichtigung der Integration der Anwendungssysteme des klinischen Registers in die vorhandene Informationsinfrastruktur einer klinisch-wissenschaftlichen Versorgungseinrichtung entwickelt. Damit liegt erstmals ein umfassender allgemeiner Leitfaden für alle Entwicklungsphasen eines klinischen Registers und seiner Integrationsaspekte vor.

Der entwickelte Leitfaden konnte erfolgreich für die Planung und den Aufbau des neu am Universitätsklinikum Heidelberg einzuführenden klinischen Myelomregisters angewendet werden. Dabei konnte die Eignung von elektronischen Datenerfassungssystemen für die Realisierung

eines klinischen Registers systematisch bewertet, ein allgemeines Metadatenmanagementsystem umgesetzt und erprobt und ein automatisiertes Verfahren für die Integration strukturierter Informationen aus dem Heidelberger Krankenhausinformationssystem ausgearbeitet und evaluiert sowie die Eignung von Text Mining Verfahren für die Übernahme unstrukturierter Informationen aus Arztbriefen untersucht werden. Die dabei entstandenen Methoden und Werkzeuge sind größtenteils auch für den Aufbau anderer klinischer Register nutzbar.

Es kann davon ausgegangen werden, dass klinische Register auch in Zukunft eine wichtige Methode der klinischen Forschung sein werden. Mit dem entwickelten Leitfaden wurde ein Werkzeug geschaffen, das Nutzen und Nachhaltigkeit klinischer Register durch systematische Planung und Integration erhöhen kann.

LITERATUR

- Adametz H u. Billig A (2009)
Semantische Interoperabilität: Semantische Konflikte.
Fraunhofer-Institut für Software- und Systemtechnik ISST White Paper 2.
- Athey BD, Braxenthaler M, Haas M u. Guo Y (2013)
transSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research.
AMIA Jt Summits Transl Sci Proc 2013: 6–8.
- Aulitzky W u. Becker G (2011)
Aufbau klinischer Krebsregister.
Onkologe (Berl) 17.2: 115–120.
- Binder H u. Blettner M (2015)
Big data in medical science—a biostatistical view.
Dtsch Arztebl Int 112.9: 137–142.
- Brooke EM (1974)
The current and future use of registers in health information systems.
Bd. 8. WHO, Geneva.
- Brühl T (2016)
Hinweise zum Recherchieren und Beschaffen wissenschaftlicher Literatur.
[Online im Internet:] <http://www.fbo3.uni-frankfurt.de/46036789/literaturrecherche.pdf>
[Stand: 29.10.2016, 11:25].
- Bundesministerium für Gesundheit (Hrsg.):
Ziele des Nationalen Krebsplans.
- Camelot Consulting (2012)
The Link King: Record Linkage and Consolidation Software.
[Online im Internet:] <http://www.the-link-king.com/>
[Stand: 7.11.2016, 13:17].
- Capurro D, Barbe M, Daza C, Maria JS, Trincado J u. Gomez I (2015)
ClinicalTime: Identification of Patients with Acute Kidney Injury using Temporal Abstractions and Temporal Pattern Matching.
AMIA Jt Summits Transl Sci Proc 2015: 46–50.
- Chang CC u. Lin CJ (2011)
LIBSVM: A library for support vector machines.
ACM Trans Intell Syst Technol 2.3: 1–27.
- Christen P (2012)
Data matching: Concepts and techniques for record linkage, entity

- resolution, and duplicate detection.*
Springer, Berlin und New York.
- Drepper J u. Semler SC (2016)
IT-Infrastrukturen in der patientenorientierten Forschung: Aktueller Stand und Handlungsbedarf - 2015.
1. Auflage. Akademische Verlagsgesellschaft AKA, Berlin.
- DWG-Gesellschaft (Hrsg.):
Was ist ein Register.
[Online im Internet:] <http://www.dwg.org/wirbelsaeulenregister/ueber-das-projekt/was-ist-ein-register/> [Stand: 10.7.2015, 11:10].
- Eckstein S (2011)
Informationsmanagement in der Systembiologie: Datenbanken, Integration, Modellierung.
Springer, Berlin und New York.
- European Medicines Agency (Hrsg.):
Guideline for Good Clinical Practice. Media, PA.
- Feinstein AR (1977)
Clinical biostatistics.
C.V. Mosby, Saint Louis.
- Firnborn D, Ganzinger M, Muley T, Thomas M u. Knaup P (2015)
A Generic Data Harmonization Process for Cross-linked Research and Network Interaction. Construction and Application for the Lung Cancer Phenotype Database of the German Center for Lung Research.
Methods Inf Med 54.5: 455–460.
- Ganslandt T, Mate S, Helbing K, Sax U u. Prokosch HU (2011)
Unlocking Data for Clinical Research - The German i2b2 Experience.
Appl Clin Inform 2.1: 116–127.
- Gliklich RE u. Dreyer NA (2007)
Registries for evaluating patient outcomes: A user's guide.
1. Aufl. Agency for Healthcare Research and Quality, Rockville, MD.
- Gliklich RE u. Dreyer NA (2010)
Registries for evaluating patient outcomes: A user's guide.
2. Aufl. Agency for Healthcare Research and Quality, Rockville, MD.
- Gliklich RE u. Dreyer NA (2014)
Registries for evaluating patient outcomes: A user's guide: Volume 1.
3. Aufl. Agency for Healthcare Research and Quality, Rockville, MD.
- Hemkens LG, Contopoulos-Ioannidis DG u. Ioannidis JPA (2016)
Current use of routinely collected health data to complement randomized controlled trials: a meta-epidemiological survey.
CMAJ open 4.2: E132–40.
- Herschel M (2010)
Datenintegration & Datenherkunft: Verteilung, Autonomie und

- Heterogenität*. Tübingen.
 [Online im Internet:] http://db.inf.uni-tuebingen.de/staticfiles/teaching/ws1011/integration/o3_matvirtvergleich.pdf
 [Stand: 16.7.2015, 10:35].
- Huber S (2014)
Informationsintegration in dynamischen Unternehmensnetzwerken: Architektur, methode und anwendung.
 Springer Gabler, Wiesbaden.
- International Conference on Harmonisation (Hrsg.):
Structure and Content of Clinical Study Reports. Ottawa.
 [Online im Internet:] <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073113.pdf> [Stand: 7.11.2016, 14:41].
- International Society for Pharmaceutical Engineering (2008)
Guidelines for good pharmacoepidemiology practices (GPP).
 Bd. 17.
- Jones J (2014)
 Entwicklung eines Metadatenmanagementsystems für das Multiple Myelom in Anlehnung an ISO/IEC 11179.
 Diplomarbeit. Heidelberg und Heilbronn: Universität Heidelberg und Hochschule Heilbronn.
- Jurczyk P, Lu JJ, Xiong L, Cragan JD u. Correa A (2011)
FRIL: Fine-Grained Records Integration and Linkage Tool.
 [Online im Internet:] <http://fril.sourceforge.net/>
 [Stand: 8.11.2016, 11:19].
- Kaspar M, Ertl M, Fette G, Dietrich G, Toepfer M, Angermann C, Stork S u. Puppe F (2016)
 Data Linkage from Clinical to Study Databases via an R Data Warehouse User Interface. Experiences from a Large Clinical Follow-up Study.
Methods Inf Med 55.4: 381–386.
- Kathmann W (2008)
 Krebsregister: Was können und sollen sie leisten?
Dtsch Med Wochenschr 133.40: p33.
- Klinkhammer-Schalke M, Marschner N u. Hofstädter F (2012)
 Register in der onkologischen Versorgungsforschung.
Onkologe (Berl) 18.2: 142–150.
- Krauss F (2015)
 Informationsextraktion aus klinischen Texten.
 Bachelorarbeit. Heidelberg: Universität Heidelberg.
- Kreuzthaler M, Schulz S u. Berghold A (2015)
 Secondary use of electronic health records for building cohort studies through top-down information extraction.
J Biomed Inform 53: 188–195.

- Lehmann TM (2005)
Handbuch der Medizinischen Informatik.
 2., vollst. neu bearb. Aufl. Hanser, München und Wien.
- Leiner F (1994)
 Systematische Planung Klinischer Dokumentation.
 Dissertation. Heidelberg: Ruprecht-Karls-Universität.
- Leiner F u. Haux R (1996)
 Systematic planning of clinical documentation.
Methods Inf Med 35.1: 25–34.
- Leiner F, Gaus W, Haux R, Knaup-Gregori P, Pfeiffer KP u. Wagner J (2011)
Medizinische Dokumentation: Lehrbuch und Leitfaden Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung.
 6. Aufl. Schattauer GmbH Verlag für Medizin und Naturwissenschaften, s.l.
- Leser U u. Naumann F (2007)
Informationsintegration: Architekturen und Methoden zur Integration verteilter und heterogener Datenquellen.
 1. Aufl. Dpunkt-Verl., Heidelberg.
- Lindberg DA, Humphreys BL u. McCray AT (1993)
 The Unified Medical Language System.
Methods Inf Med 32.4: 281–291.
- Lindörfer D u. Mansmann U (2014)
 A proposal of Checklist Items for evaluating a Patient Registry Software System, 149.
 In: Rienhoff O (Hrsg.): *Big Data und Forschungsinfrastruktur.* German Medical Science GMS Publishing House, Düsseldorf.
- Lindörfer D u. Mansmann U (2015)
 A Comprehensive Assessment Tool for Patient Registry Software Systems: The CIPROS Checklist.
Methods Inf Med 54.5: 447–454.
- Löbe M, Kropf S, Meineke F, Reißner F u. Brunkhorst F (2013)
 OpenClinica als zentraler Baustein eines Integrierten Forschungs- und Behandlungszentrums, 277.
 In: Handels H (Hrsg.): *Im Focus das Leben.* German Medical Science GMS Publishing House, Düsseldorf.
- Löprrich M, Ganzinger M, Goldschmidt H u. Knaup-Gregori P (2014a)
 Ein flexibles und erweiterbares Record Linkage Framework auf Basis von Talend Open Studio, 221.
 In: Rienhoff O (Hrsg.): *Big Data und Forschungsinfrastruktur.* German Medical Science GMS Publishing House, Düsseldorf.
- Lorenz W (1981)
 53. Eine Einführung in die Terminologie und Problematik.
Langenbecks Arch Chir 355.1: 387–391.

- Lusti M (2002)
Data warehousing und data mining: Eine Einführung in entscheidungsunterstützende Systeme.
 2., überarb. und erw. Aufl. Springer, Berlin und Heidelberg.
- Marschner N (2009)
 Krebsregister in der ambulanten Versorgung: Aufwand und Nutzen.
Forum 24: 35–38.
- Mathis S u. Wild C (2008)
Register für klinische und gesundheitsökonomische Fragestellungen: Einsatzbereiche von kardiovaskulären, wirbelsäulenspezifischen und neurologischen Registern und Good Practice Strategien für die Arbeit mit Registern: HTA-Projektbericht. Hrsg. von Ludwig Blotzmann Institut. Wien.
 [Online im Internet:] <http://eprints.hta.lbg.ac.at/788/>
 [Stand: 5.4.2016, 11:55].
- Maur E u. Winter R (2003)
Data Warehouse Management: Das St. Galler Konzept zur ganzheitlichen Gestaltung der Informationslogistik.
 Springer Berlin Heidelberg, Imprint und Springer, Berlin, Heidelberg.
- McCallum A (2002)
MALLET: A Machine Learning for Language Toolkit.
 [Online im Internet:] <http://mallet.cs.umass.edu>
 [Stand: 25.10.2016, 09:19].
- Michalik C, Dress J, Ngouongo S, Staubert S, Weber U, Brockmeyer N, Paulus U u. Stausberg J (2014)
 Requirements and tasks of cohorts and registers, the German KoRegIT project.
Stud Health Technol Inform 205: 1085–1089.
- Min JK, Dunning A, Lin FY, Achenbach S, Al-Mallah MH, Berman DS, Budoff MJ, Cademartiri F, Callister TQ, Chang HJ, Cheng V, Chinnaiyan KM, Chow B, Delago A, Hadamitzky M, Hausleiter J, Karlsberg RP, Kaufmann P, Maffei E, Nasir K, Pencina MJ, Raff GL, Shaw LJ u. Villines TC (2011)
 Rationale and design of the CONFIRM (COronary CT Angiography EvaluationN For Clinical Outcomes: An InteRnational Multicenter) Registry.
J Cardiovasc Comput Tomogr 5.2: 84–92.
- Morton T, Kottmann J, Baldrige J u. Bierner G (2005)
OpenNLP: A Java-based NLP Toolkit.
 [Online im Internet:] <http://opennlp.sourceforge.net>
 [Stand: 25.10.2016, 09:30].
- Müller D, Augustin M, Banik N, Baumann W, Bestehorn K, Kieschke J, Lefering R, Maier B, Mathis S, Rustenbach SJ, Sauerland S,

- Semler SC, Stausberg J, Sturm H, Unger C u. Neugebauer EAM (2010)
 Memorandum Register für die Versorgungsforschung:
 Memorandum registry for health services research.
Gesundheitswesen 72.11: 824–839.
- Murphy SN, Mendis M, Hackett K, Kuttan R, Pan W, Phillips LC, Gainer V, Berkowicz D, Glaser JP, Kohane I u. Chueh HC (2007)
 Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside.
AMIA Annu Symp Proc: 548–552.
- Németh L (2014)
Hunspell.
 [Online im Internet:] <http://hunspell.github.io/>
 [Stand: 25.10.2016, 09:25].
- Nonnemacher M, Weiland D u. Stausberg J (2007)
Datenqualität in der medizinischen Forschung: Leitlinie zum adaptiven Management von Datenqualität in Kohortenstudien und Registern.
 Bd. 4. Med.-Wiss. Verl.-Ges, Berlin.
- Ogden CK, Postgate JP u. Richards IA (1923)
The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism.
 London.
- Pang X, Kozlowski N, Wu S, Jiang M, Huang Y, Mao P, Liu X, He W, Huang C, Li Y u. Zhang H (2014)
 Construction and management of ARDS/sepsis registry with REDCap.
J Thorac Dis 6.9: 1293–1299.
- Pentaho Corporation (2016)
Data Integration.
 [Online im Internet:] <http://www.pentaho.com>
 [Stand: 27.10.2016, 11:22].
- Plotnicki L, Kohl CD, Höcker B, Krupka K, Rahmel A, Pape L, Hoyer P, Marks SD, Webb NJA, Söylemezoglu O, Topaloglu R, Szabo AJ, Seeman T, Marlies Cornelissen EA, Knops N, Grenda R u. Tönshoff B (2013)
 The CERTAIN Registry: a novel, web-based registry and research platform for pediatric renal transplantation in Europe.
Transplant Proc 45.4: 1414–1417.
- Pommerening K u. Muscholl M (2015)
 Datenschutz und IT-Sicherheit in der Medizin, 385–420.
 In: Dickhaus H u. Knaup-Gregori P (Hrsg.): *Biomedizinische Technik - Medizinische Informatik*.

- Püschmann H, Haferkamp G, Scheppokat KD, Vinz H u. Wegner M (2006)
Vollständigkeit und Qualität der ärztlichen Dokumentation in Krankenakten: Untersuchung zu Krankenunterlagen aus Chirurgie, Orthopädie, Innerer Medizin und Neurologie.
Dtsch Arztebl Int 103.3: A–121.
- Rüping S u. Graf N (2013)
Datenanalyse: Big Data in der Medizin.
Dtsch Arztebl Int 110.41: A–1926.
- Rustenbach SJ, Heyer K, Reppenhagen K u. Augustin M (2011)
Registerforschung in der Dermatologie.
MMW Munch Med Wochenschr 62.3: 189–195.
- Singh SK, Malik A, Firoz A u. Jha V (2012)
CDKD: a clinical database of kidney diseases.
BMC Nephrol 13.23.
- Talend (2016)
Talend Open Studio for Data Integration.
[Online im Internet:] <http://de.talend.com> [Stand: 27.10.2016, 11:20].
- Torres V (2016)
Assessment of Exported Clinical Data from Hospital Information System for Clinical Reuse in Multiple Myeloma Research.
Magister Thesis. Santiago de Chile: Universidad de Chile.
- Trotter JP (2002)
Patient registries: a new gold standard for "real world" research.
Ochsner J 4.4: 211–214.
- Verband der forschenden Arzneimittelhersteller (Hrsg.):
Register auf dem Vormarsch: Datensammlungen in der Medizin.
[Online im Internet:] <https://www.vfa-patientenportal.de/patienten-und-versorgung/register-auf-dem-vormarsch.html>
[Stand: 7.11.2016, 11:59].
- Viviani L, Zolin A, Mehta A u. Olesen HV (2014)
The European Cystic Fibrosis Society Patient Registry: valuable lessons learned on how to sustain a disease registry.
Orphanet J Rare Dis 9: 81.
- Wegscheider K (2004)
Medizinische Register: Nutzen und Grenzen.
Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 47.5: 416–421.
- Wöhe G u. Döring U (2013)
Einführung in die allgemeine Betriebswirtschaftslehre.
25., überarb. und aktualisierte Aufl. Vahlen, München.
- Yoder RD, Swearingen DR, Schenthal JE, Sweeney JW u. Nettleton JR WJ (1964)

An Automated Clinical Information System.

Methods Inf Med 3.2: 45–50.

Zylka-Menhorn V (2015)

Klinische Krebsregister: Zur Halbzeit noch eine Großbaustelle.

Dtsch Arztebl Int 112.44: A–1818.

EIGENE VERÖFFENTLICHUNGEN

ZEITSCHRIFTENAUFsätze

Gietzelt M, **Löpprich M**, Karmen C, Knaup P u. Ganzinger M (2016b)
Models and Data Sources Used in Systems Medicine. A Systematic Literature Review.

Methods Inf Med 55.2.

Löpprich M, Krauss F, Ganzinger M, Senghas K, Riezler S u. Knaup P (2016)

Automated Classification of Selected Data Elements from Free-text Diagnostic Reports for Clinical Research.

Methods Inf Med 55.4: 373–380.

Reuner G, Fields AC, Wittke A, **Löpprich M** u. Pietz J (2013)

Comparison of the developmental tests Bayley-III and Bayley-II in 7-month-old infants born preterm.

Eur J Pediatr 172.3: 393–400.

WEITERE MEDLINE-GEFÜHRTE ORIGINALARBEITEN

Gietzelt M, Hofer T, Knaup-Gregori P, König R, **Löpprich M**, POOS A u. Ganzinger M (2016a)

The Use of Tools, Modelling Methods, Data Types, and Endpoints in Systems Medicine: A Survey on Projects of the German e:Med-Programme.

Stud Health Technol Inform 228: 670–674.

Löpprich M, Jones J, Meinecke MC, Goldschmidt H u. Knaup P (2014b)

A reference data model of a metadata registry preserving semantics and representations of data elements.

Stud Health Technol Inform 205: 368–372.

Löpprich M u. Knaup P (2016)

Analysis of Existing Guidelines for the Systematic Planning Process of Clinical Registries.

Stud Health Technol Inform 228: 451–455.

Torres V, Cerda M, Knaup P u. **Löpprich M** (2016)

Assessment of Automatically Exported Clinical Data from a Hospital Information System for Clinical Research in Multiple Myeloma.

Stud Health Technol Inform 228: 332–336.

WEITERE KONFERENZBEITRÄGE

- Firnkor D, Karmen C, Kohl CD, **Löpprich M**, Knaup-Gregori P u. Ganzinger M (2014)
An Ontology-Generator for izb2, 1.
In: Prokosch HU, Dugas M u. Semler S (Hrsg.): *GMDS/IMIA Workshop "Research Databases" & 2nd European izb2 Academic User Group Meeting*.
- Ganzinger M, Gietzelt M, Karmen C, **Löpprich M** u. Knaup P (2014)
Datenintegration und IT-Werkzeuge für die Systemmedizin am Beispiel von CLIOMMICS, 3.
In: Löhnhardt L, Sax U u. Beissbarth T (Hrsg.): *Integration und Visualisierung von Analyseergebnissen in der translationalen Forschung*.
- Gietzelt M, **Löpprich M**, Karmen C, Knaup-Gregori P u. Ganzinger M (2015)
Aktuelle Methoden in der Systemmedizin – Ein systematischer Literaturreview, 185.
In: Breil B (Hrsg.): *Fortschritt durch Forschung*. German Medical Science GMS Publishing House, Düsseldorf.
- Haas EM, Mai EK, **Löpprich M**, Lücke S, Kunz C, Pritsch M, Knaup-Gregori P, Hillengass J u. Goldschmidt H (2015)
Systematic classification of death causes in multiple myeloma patients treated with high-dose therapy.
Clin Lymphoma Myeloma Leuk 15: e62.
- Krauss F, Senghas K, Knaup-Gregori P u. **Löpprich M** (2015)
Eignung computerlinguistischer Methoden zur Informationsextraktion aus Arztbriefen, 253.
In: Breil B (Hrsg.): *Fortschritt durch Forschung*. German Medical Science GMS Publishing House, Düsseldorf.
- Löpprich M**, Ganzinger M, Goldschmidt H u. Knaup-Gregori P (2014a)
Ein flexibles und erweiterbares Record Linkage Framework auf Basis von Talend Open Studio, 221.
In: Rienhoff O (Hrsg.): *Big Data und Forschungsinfrastruktur*. German Medical Science GMS Publishing House, Düsseldorf.
- Löpprich M**, Stausberg J u. Knaup-Gregori P (2013)
Inhalt und Aufbau eines Registerprotokolls, 148.
In: Handels H (Hrsg.): *Im Focus das Leben*. German Medical Science GMS Publishing House, Düsseldorf.
- Senghas K, **Löpprich M** u. Knaup-Gregori P (2015)
Visuelle Integration einer Patientenliste mit OpenClinica, 079.
In: Breil B (Hrsg.): *Fortschritt durch Forschung*. German Medical Science GMS Publishing House, Düsseldorf.

ANHANG



TESTPROTOKOLL DYNAMISCHES DOKUMENTATIONSMODELL

Zur Prüfung, ob sich ein spezifisches Electronic Data Capture (EDC)-System zur Umsetzung eines dynamischen Dokumentationsmodells eignet, kann folgendes Testprotokoll implementiert werden:

Es sind drei eCRFs zu entwerfen: *Diagnose*, *Therapie* und *Labor*. Sind Events zu definieren, so können diese identisch bezeichnet werden. Jedem Event wird ein eCRF zugeordnet. Die Events oder eCRFs, falls eine Eventdefinition nicht erforderlich ist, sollen variabel kombiniert werden können beispielsweise *Diagnose-Therapie-Labor* oder *Diagnose-Labor-Therapie*. Zusätzlich sollen Events oder eCRFs variabel wiederholbar sein beispielsweise *Diagnose-Diagnose-Therapie* oder *Diagnose-Therapie-Therapie*. Durch diese beiden Eigenschaften, variabel kombinierbar und variabel wiederholbar, sind individuelle Krankheits- und Behandlungsverläufe möglich. Eine Instanz kann damit den Verlauf *Diagnose-Therapie-Labor* haben, eine andere *Diagnose-Labor-Therapie-Labor-Diagnose-usw.*

Auf jedem eCRF sind mehrere Merkmale mit eindeutigem Variablenamen enthalten. Als Datentypen sind Ganz- und Gleitkommazahlen, Einfach- und Mehrfachauswahllisten, Text- und Datumsfelder vorgesehen. Ist ein Merkmal nicht verfügbar (NA; not available) oder unbekannt (UNK; unknown), dann sollte dies ebenfalls erfasst werden können: entweder durch Systemfunktionen, zusätzliche Codierungen oder, falls technisch nicht anders realisierbar, durch eine Dokumentationsanleitung beispielsweise, dass unbekannte Zahlenwerte mit -1 dokumentiert werden.

Die Konsistenz der Merkmale wird über Plausibilitätsregeln sichergestellt. Werden diese verletzt, soll der Benutzer eine Fehlermeldung oder einen Warnhinweis erhalten. Bei Fehlermeldungen ist eine Korrektur verpflichtend, bei Warnhinweisen optional. Eine Echtzeitprüfung merkmalsbezogener Regeln ist wünschenswert. Weitere Plausibilitäten sind zwischen Merkmalen eines eCRFs und zwischen Merkmalen auf verschiedenen eCRFs vorgesehen. Die Prüfung eCRF-bezogener und eCRF-übergreifender Plausibilitäten kann bei Speicherung des eCRFs oder durch ein spezifisches Validierungsmodul erfolgen.

Merkmale sind in Abhängigkeit eines anderen Merkmals auf der Benutzeroberfläche anzuzeigen oder auszublenden. Auch eine feste oder variable Anzahl Wiederholungen desselben Merkmals oder der Merkmalsgruppe sollte realisiert werden können.

Tabelle 18: Test- und Implementierungsprotokoll zur Umsetzung eines dynamischen Dokumentationsmodells.

DIAGNOSE	
Diagnosedatum	
Variablenname	diag_date
Datentyp	Datum (tt.mm.yyyy)
Bemerkung	Wenn Tag unbekannt, dann Eingabe von mm.yyyy. Wenn Tag und Monat unbekannt, dann yyyy. Wenn Datum vollständig unbekannt, dann UNK (unknown). Wenn Datum vollständig nicht verfügbar, dann NA (not available). Alternativ: Eingabe von 01.mm.yyyy, 01.01.yyyy bzw. 01.01.1900.
Diagnose	
Variablenname	diag
Datentyp	Ganzkommazahl, Einfachauswahl
Wertemenge (Code, Wert)	0, auswählen 1, Diagnose 1 2, Diagnose 2
Bemerkung	Standardwert ist 0. Wenn Diagnose unbekannt, dann UNK. Wenn Diagnose nicht verfügbar, dann NA. Alternativ: zusätzliche Werte {-1, UNK; -2, NA}.
THERAPIE	
Therapietyp	
Variablenname	ther_type
Datentyp	Ganzkommazahl, Einfachauswahl
Wertemenge (Code, Wert)	0, auswählen 1, Therapie 1 2, Therapie 2
Bemerkung	siehe Diagnose (diag)
Startdatum	
Variablenname	ther_start
Datentyp	Datum (tt.mm.yyyy)
Bemerkung	siehe Diagnosedatum (diag_date)
Enddatum	
Variablenname	ther_end
Datentyp	Datum (tt.mm.yyyy)
Plausibilitäten	<ul style="list-style-type: none"> • Startdatum (ther_start) muss vor Enddatum liegen • Datum darf nur eingegeben werden, wenn Therapietyp (ther_type) gleich 1
Bemerkung	siehe Diagnosedatum (diag_date)

Tabelle 18: (Fortsetzung)

Dauer in Monaten	
Variablenname	ther_duration
Datentyp	Ganzkommazahl
Plausibilitäten	<ul style="list-style-type: none"> • Dauer darf nur eingegeben werden, wenn Therapietyp (ther_type) gleich 2 • Wertebereich 0 bis 24; Fehlermeldung bei unterschreiten; Warnhinweis bei überschreiten
Zyklen	
Variablenname	ther_cyc
Datentyp	Ganzkommazahl
Plausibilität	Wertebereich 0 bis 5; Fehlermeldung bei unter- und überschreiten
Bemerkung	Merkmale Medikament Zyklus X (ther_cyc_drug_X) und ggfs. sonstiges Medikament Zyklus X (ther_cyc_drug_other_X) erscheinen so oft wie Wert, der gewählt wurde.
Medikament Zyklus X	
Variablenname	ther_cyc_drug_X
Datentyp	Ganzkommazahl, Mehrfachauswahl, wiederholbar
Wertemenge (Code, Wert)	0, auswählen 1, Medikament 1 2, Medikament 2 9, sonstiges
Plausibilitäten	<ul style="list-style-type: none"> • Medikament Zyklus X darf nur eingegeben werden, wenn Zyklen (ther_cyc) größer 0 • Medikament Zyklus X darf so häufig eingegeben werden, wie Wert Zyklen (ther_cyc) • Wenn Medikament Zyklus X gleich 9, dann erscheint sonstiges Medikament Zyklus X (ther_cyc_drug_other_X)
Bemerkung	siehe Diagnose (diag)
sonstiges Medikament Zyklus X	
Variablenname	ther_cyc_drug_other_X
Datentyp	Text, wiederholbar
Plausibilitäten	<ul style="list-style-type: none"> • sonstiges Medikament Zyklus X darf nur eingegeben werden, wenn Zyklen (ther_cyc) größer 0 • sonstiges Medikament Zyklus X darf so häufig eingegeben werden, wie Wert Zyklen (ther_cyc) • sonstiges Medikament Zyklus X darf nur eingegeben werden, wenn Medikament Zyklus X (ther_cyc_drug_X) gleich 9

Tabelle 18: (Fortsetzung)

LABOR	
vorhanden	
Variablenname	lab_avail
Datentyp	Ganzkommazahl, Einfachauswahl
Wertemenge (Code, Wert)	0, auswählen 1, ja 2, nein
Plausibilität	Merkmale Datum (lab_date), Wert 1 (lab_val1), Wert 1 Einheit (lab_val1_unit) erscheinen nur, wenn vorhanden gleich 1
Bemerkung	siehe Diagnose (diag)
Labordatum	
Variablenname	lab_date
Datentyp	Datum (tt.mm.yyyy)
Plausibilität	Labordatum darf nur eingegeben werden, wenn vorhanden (lab_avail) gleich 1 ist
Bemerkung	siehe Diagnosedatum (diag_date)
Wert 1	
Variablenname	lab_val1
Datentyp	Ganzkommazahl
Plausibilitäten	<ul style="list-style-type: none"> • Wert 1 darf nur eingegeben werden, wenn vorhanden (lab_avail) gleich 1 ist • Wertebereich 0 bis 10, wenn Wert 1 Einheit (lab_val1_unit) gleich 1; Fehlermeldung bei unterschreiten; Warnhinweis bei überschreiten • Wertebereich 100 bis 1000, wenn Wert 1 Einheit (lab_val1_unit) gleich 2; Warnhinweis bei unter- und überschreiten
Bemerkung	Wenn Wert 1 unbekannt, dann UNK. Wenn Wert 1 nicht verfügbar, dann NA. Alternativ: negative Zahlen {-1, UNK; -2, NA}.
Wert 1 Einheit	
Variablenname	lab_val1_unit
Datentyp	Ganzkommazahl, Einfachauswahl
Wertemenge (Code, Wert)	0, auswählen 1, Einheit 1 2, Einheit 2
Plausibilitäten	<ul style="list-style-type: none"> • Wert 1 Einheit darf nur eingegeben werden, wenn vorhanden (lab_avail) gleich 1 ist • Wenn Wert 1 Einheit gleich 1, dann Wert 1 zwischen 0 und 10 • Wenn Wert 1 Einheit gleich 2, dann Wert 1 zwischen 100 und 1000

Tabelle 18: (Fortsetzung)

Wert 2	
Variablenname	lab_val2
Datentyp	Gleitkommazahl
Plausibilitäten	<ul style="list-style-type: none"> • Wert 2 darf nur eingegeben werden, wenn vorhanden (lab_avail) gleich 1 ist • Wertebereich 0 bis 5, wenn unmittelbar vorausgegangener Therapietyp (ther_type) gleich 1; Fehlermeldung bei unterschreiten; Warnhinweis bei überschreiten • Wertebereich 10 bis 15, wenn unmittelbar vorausgegangener Therapietyp (ther_type) gleich 2; Warnhinweis bei unter- und unterschreiten
Bemerkung	siehe Wert 1 (lab_val1)

Ist die Implementierung der eCRFs möglich, können folgende beiden Instanzen mit individuellen Krankheits- und Therapieverläufen aufgenommen werden.

Tabelle 19: Testinstanz 1 für das Test- und Implementierungsprotokoll zur Umsetzung eines dynamischen Dokumentationsmodells aus Tabelle 18.

PATIENT 1	
Diagnose	
Diagnosedatum	01.01.2016
Diagnose	1
Therapie	
Therapietyp	1
Startdatum	01.02.2016
Enddatum	05.02.2016
Zyklen	1
Medikament Zyklus 1	1, 2
Labor	
vorhanden	1
Labordatum	03.02.2016
Wert 1	5
Wert 1 Einheit	1
Wert 2	3-5

Tabelle 19: (Fortsetzung)

Therapie	
Therapietyp	2
Startdatum	10.02.2016
Dauer in Monaten	4
Zyklen	2
Medikament Zyklus 1	1, 2
Medikament Zyklus 2	1, 9
sonstiges Medikament Zyklus 2	„Medikament XY“
Medikament Zyklus 3	2, 9
sonstiges Medikament Zyklus 3	„Medikament XY“

Tabelle 20: Testinstanz 2 für das Test- und Implementierungsprotokoll zur Umsetzung eines dynamischen Dokumentationsmodells aus Tabelle 18.

PATIENT 2	
Diagnose	
Diagnosedatum	01.06.2016
Diagnose	2
Labor	
vorhanden	1
Labordatum	03.06.2016
Wert 1	500
Wert 1 Einheit	2
Wert 2	29
Therapie	
Therapietyp	2
Startdatum	10.02.2016
Dauer in Monaten	4
Zyklen	1
Medikament Zyklus 1	9
sonstiges Medikament Zyklus 1	„Medikament XY“

B

SYSTEMBEWERTUNG OPENCLINICA UND ONKOSTAR

Tabelle 21: Systembewertung der elektronischen Dokumentationssysteme *OpenClinica* und *Onkostar*. Die Bewertung erfolgte im April 2015 mit den zum damaligen Zeitpunkt aktuellen und verfügbaren Versionen.

SYSTEM	OPENCLINICA	ONKOSTAR
Hersteller	OpenClinica, LLC Waltham, MA 02451 USA	IT-Choice, Software AG 76189 Karlsruhe Deutschland
Entwickelt seit	2005	2012
Im Produktiveinsatz	ja mehrere nationale und internationale Projekte	ja aktuell ein Projekt am Universitätsklinikum Heidelberg
Lizenz	OpenSource LGPL kommerzielle Enterprise Version, kostenlose Community Version	unbekannt
Kosten	keine (Community Version)	unbekannt
Weiterentwicklung	ja	ja
1. GENERELLE EIGENSCHAFTEN		
Multi-Projekttauglichkeit	ja mehrere Projekte; jedes Projekt mit eigenem Patientenstamm, eigenen Formularen und Zugriffsberechtigungen	nein verschiedene Patientenstämme; keine unterschiedlichen Formulare pro Patientenstamm
Software-aktualisierung	ja	unbekannt
technische Dokumentation	vorhanden	aktuell nicht vorhanden
2. ARCHITECTURELLE GESICHTSPUNKTE		
Systemarchitektur	Java Webapplikation; PostgreSQL	Java J2EE Webapplikation; MySQL
Laufzeitumgebung	Apache Tomcat	Apache Tomcat
Skalierungsfähigkeit	keine Bedenken mehrere Projekte; mehrere Zentren pro Projekt; Patienten mit individuellem Krankheitsverlauf	keine Bedenken ein Projekt; Patienten mit individuellem Krankheitsverlauf
Quellcode-anpassungen	möglich	nicht möglich

Tabelle 21: (Fortsetzung)

3. MENSCH-MASCHINE SCHNITTSTELLEN		
Anwendungs- oberfläche	Zugriff über Webbrowser; Menüeinträge Englisch; eCRFs und Merkmale Deutsch möglich	Zugriff über Webbrowser; Menüeinträge Deutsch; eCRFs und Merkmale Deutsch möglich
Startseite	Übersicht mit Patientenanzahl, abgeschlossener / nicht abgeschlossener Events, zugewiesener Aufgaben	Übersicht mit zuletzt bearbeiteten Patienten, zugewiesener Aufgaben, Terminen
Zusatzdienste	vorhanden Diskussionsforen, Webseiten	nicht vorhanden
4. FUNKTIONEN DATENMANAGEMENT		
eCRF Definition	sehr umfangreich in Excel; erfordert Einarbeitung; weitere eCRFs aus Templates ableitbar	umfangreich über Webeditor; interne Restriktionen schränken Definition ein; erfordert Einarbeitung; Standard-Kataloge und eCRFs verfügbar; jedes eCRF separat zu erstellen
eCRF Versionierung	vorhanden an klinische Studien orientiert; jede Änderung ergibt neue eCRF-Version (Entwicklungsumgebung zweckmäßig); alte eCRF-Versionen nicht löschtbar; eCRF-Versionsauswahl pro Patient	nicht testbar
Ungeplante Visiten	vorhanden individuelle Krankheits- und Therapieverläufe möglich	vorhanden individuelle Krankheits- und Therapieverläufe möglich
Grafischer Editor	nicht vorhanden Definition in Excel; Entwicklung aufwändig	vorhanden Entwicklung verschachtelt
Pseudonymisierung	vorhanden	nicht vorhanden Klartextangaben bei Patientenaufnahme gefordert
5. FUNKTIONEN DATENQUALITÄT		
Plausibilitäts- prüfungen	sehr umfangreich merkmalsbezogene Regeln bei eCRF Definition in Excel; eCRF-bezogene und eCRF-übergreifende Regel in XML; erfordert Einarbeitung; grafische Übersicht aller Regeln vorhanden; grafische Übersicht aller Konflikte vorhanden	nicht testbar grafische Übersicht aller Regeln nicht vorhanden; grafische Übersicht aller Konflikte nicht vorhanden

Tabelle 21: (Fortsetzung)

Konfliktbehandlung	sehr umfangreich Fehlermeldung in Echtzeit und bei Speichern des eCRF; nicht behobene Fehler werden protokolliert; Fehler können Anwendern zugewiesen werden	mäßig Fehlermeldungen in Echtzeit und bei Speichern des eCRF
Datenexport	sehr umfangreich grafischer Editor vorhanden; Auswahl von Merkmalen möglich; verschiedene Export-Formate	sehr umfangreich grafischer Editor vorhanden; Auswahl von Merkmalen möglich; Export als CSV-Datei
6. FUNKTIONEN DATENANALYSE		
Statistische Auswertungen	mäßig nur deskriptiv	nicht vorhanden
Graphische Aufbereitung	nicht vorhanden	nicht vorhanden
7. SICHERHEITSASPEKTE		
Benutzer- autorisierung	über Benutzername und Passwort	über Benutzername und Passwort
Benutzerrollen	vorhanden Rollen vorgegeben; keine eigenen Rollen möglich	vorhanden freie Rollendefinition möglich
Audit-Trail	sehr umfangreich Benutzeraktionen werden pro Patient gespeichert; Protokollierung von Datum, neuen Wert, alten Wert	umfangreich Benutzeraktionen werden pro Event gespeichert; bei Eventlöschung auch protokollierte Benutzeraktionen entfernt
Verschlüsselung	vorhanden	vorhanden
8. QUALITATIVE BEWERTUNG DES SYSTEMS		
Benutzer- freundlichkeit	gering unübersichtliche Menüführung	sehr gut übersichtliche Menüführung
Benutzer- rückmeldungen	gering Meldungen teilweise versteckt	gut
9. BENUTZERTRAINING		
Trainings- materialien	vorhanden überwiegend Englisch	nicht vorhanden
Komplexität der Arbeitsflüsse	gering Aufnahme neuer Patienten einfach möglich; Events in beliebiger Reihenfolge möglich; Anzeige, dass Erfassung abgeschlossen ist	gering Aufnahme neuer Patienten einfach möglich; Events in beliebiger Reihenfolge möglich; keine Anzeige, dass Erfassung abgeschlossen ist

Tabelle 21: (Fortsetzung)

10. ANPASSBARKEIT		
Software-schnittstellen	vorhanden Zugriff über Webservices; HL7-Schnittstellen über Webservices	vorhanden HL7-Schnittstelle
Zugriffsmöglichkeiten auf Datenbanksystem	vorhanden Dokumentation des Datenbankmodells vorhanden	nicht testbar keine Dokumentation des Datenbankmodells vorhanden
Erweiterbarkeit der Funktionalität	möglich aufgrund von Produktaktualisierungen nicht empfehlenswert	nicht möglich neue Funktionen nur über Hersteller realisierbar

LEBENS LAUF

PERSONALIEN

Name: Martin Löprrich
Geburtsjahr: 1985
Geburtsort: Hermannstadt
Familienstand: verheiratet

SCHULISCHER WERDEGANG

1991–1995 Salier-Grundschule Waiblingen
1995–1998 Salier-Realschule Waiblingen
1998–2001 GHRS-Realschule Illingen
05. Juli 2001 Mittlere Reife
2001–2004 Fritz-Erler-Wirtschaftsgymnasium Pforzheim
22. Juni 2004 Allgemeine Hochschulreife

UNIVERSITÄRER WERDEGANG

2005–2010 Studium der Medizinischen Informatik
an der Universität Heidelberg und
Hochschule Heilbronn
17. November 2010 Diplomprüfung im Studiengang
Medizinische Informatik

BERUFLICHER WERDEGANG

seit 2011 Wissenschaftlicher Mitarbeiter
Institut für Medizinische Biometrie und Informatik
Sektion Medizinische Informatik
Universität Heidelberg

DANKSAGUNG

An dieser Stelle möchte ich all jenen danken, die durch ihre fachliche und persönliche Unterstützung zum Gelingen dieser Dissertationsschrift beigetragen haben:

Mein besonderer Dank gilt Frau Professorin Dr. Petra Knaup-Gregori für die Vergabe des spannenden Themas und die sehr gute Betreuung. Danke für die vielen wertvollen Diskussionen, die Bereitschaft auch kurzfristig meine Beiträge zu lesen, für die konstruktive Kritik und Förderung, die mich vorangebracht haben und für die Hilfe in jeglicher Hinsicht, die letztendlich zum erfolgreichen Abschluss meiner Arbeit beigetragen hat.

Sowie Herrn Professor Dr. Hartmut Goldschmidt für die Möglichkeit den in dieser Dissertationsschrift entwickelten Leitfaden in der Sektion Multiples Myelom anzuwenden und praktisch umzusetzen. Vielen Dank für die Unterstützung und Förderung meiner Arbeit.

Da diese Arbeit in enger Kooperation zwischen den Sektionen Medizinische Informatik und Multiples Myelom stattfand, bin ich den Mitgliedern beider Sektionen zu großem Dank verpflichtet, besonders: Frau Dr. Maria Pritsch für die vielen fachlichen Gespräche, Ratschläge und Anmerkungen, die mich auf dem Weg zur fertigen Arbeit immer wieder neue Aspekte und Ansätze entdecken ließen; Frau Kerstin Scherbaum-Lawrenz sowie dem gesamten Dokumentationsteam für die wertvollen praktischen Erfahrungen; Herrn Dr. Matthias Ganzinger für die zahlreichen Diskussionen und für die Ratschläge in Bezug auf den goldenen Schnitt.

Ich danke zudem meinen Eltern für ihre Unterstützung sowie allen Freunden, die meine Arbeit mit großem Interesse verfolgten. Ein großes Dankeschön dabei an Lena, die unter Hochdruck die Arbeit Korrektur gelesen und ihre Anmerkungen so formuliert hat, dass ich sie direkt umsetzen konnte.

Im Besonderen gilt mein Dank aber meiner Frau Kim, die gerade in der Endphase stets liebevoll, geduldig und aufmunternd für mich da war, und für ihre unglaublich hilfreiche Unterstützung und ihr Verständnis bei der Anfertigung dieser Dissertationsschrift. Und letztendlich ein Dank an Ella für die zusätzliche kleine Motivation :)