

3 Datenelemente und kontrollierte Vokabulare

Jürgen Stausberg

Die Durchsetzung von Standards zur medizinischen Dokumentation erfordert u.a. den Aufbau eines Metadata Repositories in Deutschland. Im Rahmen des Metadata Repositories finden dann kontrollierte Vokabulare wie Terminologien und Klassifikationen ihre Verwendung. Aus Sicht der Gesundheitsversorgung wären hierfür noch nationale Eigenentwicklungen ausreichend; eine weltweit vernetzte empirische Forschung ist dagegen jetzt schon von international etablierten Terminologien und Klassifikationen abhängig.

Die ISO/IEC 11179-3 Information technology – Metadata registries (MDR) definiert Metadaten allgemein als Daten, die andere Daten definieren und beschreiben [ISO/IEC, 2013]. Im engeren Sinne handelt es sich um strukturelle Metadaten. Strukturelle Metadaten in medizinischen Dokumentationssystemen bestehen aus Definitionen von Merkmalen, Wertebereichen und Abhängigkeiten [Leiner und Haux, 1996]. Synonym zu Merkmal werden auch die Worte Datenelement, Item oder Variable verwendet. Ein Metadata Repository verwaltet Metadaten [ISO/IEC, 2013]. Es ermöglicht so die Recherche nach Metadaten, die Harmonisierung von Metadaten und die Bereitstellung von Metadatenstandards [Stausberg et al., 2009]. Synonym zu Metadata Repository wird auch von Metadata Registry gesprochen.

Die in einem Metadata Repository definierten Datenelemente werden über informationstechnische Informations- und Datenmodelle in – zunehmend rechnergestützten – Dokumentationssystemen umgesetzt [Nguouongo und

[Stausberg, 2011a]. Der Nutzer eines Dokumentationssystems findet die Datenelemente dann als Teil einer strukturierten Erfassung oder als Auswahlmöglichkeit für Recherchen und Analysen vor. Ein Metadata Repository deckt in etwa die Ebene des Merkmalskatalogs nach Leiner und Haux ab [Leiner und Haux, 1996]; es erweitert die Verwaltung des Merkmalskatalogs eines Vorhabens aber um die Abbildung und Verknüpfung vieler Merkmalskataloge vieler Vorhaben [Stausberg et al., 2009]. In anderen Ländern sind Metadata Repositories bereits auf nationaler Ebene implementiert, so für die Versorgung in Australien mit der Metadata Online Registry [METeOR, 2014] und in den USA für die Forschung mit dem Cancer Data Standards Registry and Repository [NCI, 2010]. Einschränkend ist aus Sicht der medizinischen Dokumentation allerdings festzuhalten, dass ein Metadata Repository nicht ausreicht, um volle semantische Interoperabilität zu erreichen. Zusätzlich sind u. a. Anlässe und Verfahren zur Ermittlung einer Information festzulegen. Dies könnte beispielsweise die Messung des Blutdrucks am rechten Oberarm nach Riva-Rocci im Sitzen (Verfahren) am ersten Montag eines Quartals (Anlass) sein. Alternativ zur Festlegung des Kontextes der Ermittlung von Informationen lässt sich der Kontext auch als Zusatzinformation erheben. Dies führt dann zu komplexen Informationsmodellen für Datenelemente [Stausberg und Hucklenbroich, 1994], die sich bislang jedoch nicht durchsetzen konnten.

Die Notwendigkeit einer Kontrolle verwendeter Datenelemente wurde frühzeitig von Betreibern rechnergestützter Anwendungssysteme im Krankenhaus erkannt und über so genannte Data Dictionaries gelöst. So wurde im Rahmen des Systems Health Evaluation through Logical Processing (HELP) am LDS Hospital in Salt Lake City, Utah, eine flache Liste der Datenelemente mit eindeutigen Identifikationskennzeichen geführt [Kupermann et al., 1991]. Dieser Ansatz gewann dann Anfang der 90er-Jahre des letzten Jahrtausends durch die Kopplung von Wissensbasen mit elektronischen Akten systemübergreifend an Bedeutung [Stausberg et al., 1995]. Linnarsson und Wigertz erweiterten die einzelnen Konzepte zu einem generischen Informationsmodell für Data Dictionaries [Linnarsson und Wigertz, 1989]. Später wurde das Feld systematisch aufgearbeitet [Bürkle, 2000]. Das Data Dictionary finden wir nun als Metadata Repository wieder.

Bei Datenelementen herrscht in Deutschland Wildwuchs. Für Dokumentationsverfahren Verantwortliche treffen für ihren Bereich – ohne erkennbare Abstimmung mit anderen Institutionen des Gesundheitswesens – eigene Festlegungen, die zu Inkompatibilitäten führen und damit die Mehrfachnutzung von Daten behindern, zum Teil auch ganz verhindern. Weitreichende Festlegungen finden sich z. B. bei der Kassenärztlichen Bundesvereinigung (KBV), die für ihre Verfahren im Internet eine Liste mit rund 110 Datenelementen (als Schlüsseltabellen bezeichnet) veröffentlicht [KBV, 2014]. Für Zwecke der einrichtungsübergreifenden Qualitätssicherung (EQS) nach § 137 Sozialgesetzbuch (SGB) V führt das AQUA-Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen eine Datenbank mit allen Datenelementen

der entsprechenden Verfahren. In der Fassung für 2015 enthält die Datenbank 1600 Felder, die den Datenelementen eines Metadata Repository entsprechen [SQG, 2014].

Der Wildwuchs bei den Dokumentationsverfahren der Selbstverwaltung unseres Gesundheitswesens lässt sich gut am Beispiel des Datenelements Geschlecht verdeutlichen. Zur Abrechnung einer stationären Behandlung im Rahmen der gesetzlichen Krankenversicherung wird das Datenelement Geschlecht mit den Kategorien männlich (kodiert mit „m“) und weiblich (kodiert mit „w“) verwendet [SCGB-V, 2014], zur Gestaltung des DRG-Systems hingegen das Datenelement Geschlecht mit den Kategorien männlich (kodiert mit „m“), weiblich („w“) und unbekannt („u“) [KHEntgG, 2013]. Viele Einrichtungen nehmen an beiden Verfahren teil. Wenn es denn Fälle mit unbekanntem Geschlecht geben sollte, so ist völlig offen, ob diese bei der Abrechnung als männlich oder als weiblich übermittelt werden. Die EQS verwendet gleich drei unterschiedliche Wertebereiche für Geschlecht, a) männlich (kodiert mit „1“) und weiblich („2“) im Regelfall, b) männlich („1“), weiblich („2“) und unbekannt („3“) im Leistungsbereich Dekubitus sowie c) männlich („1“), weiblich („2“) und nicht bestimmbar („3“) in der Neonatologie. Die KBV trifft mit ihrer Definition fast Variante b) der EQS, kodiert allerdings unbekannt mit „0“ und nicht mit „3“ (s. [KBV, 2014: Geschlecht]). Auch die medizinischen Fachgesellschaften nehmen eigene Festlegungen vor. Der Basisdatensatz Onkologie der Arbeitsgemeinschaft Deutscher Tumorzentren und der Gesellschaft der epidemiologischen Krebsregister in Deutschland definiert vier Ausprägungen von Geschlecht: „M“ = männlich, „W“ = weiblich, „S“ = Sonstiges/intersexuell, „U“ = unbekannt [Bundesanzeiger, 2014]. Die Deutsche Interdisziplinäre Vereinigung für Intensiv- und Notfallmedizin führt dann noch im Kerndatensatz Notaufnahme mit „Patientengeschlecht“ eine neue Bezeichnung ein [DIVI, 2010]. Insgesamt liegen damit bei den angeführten Beispielen bereits acht (!) unterschiedliche Definitionen von Geschlecht mit Abweichungen bei der Bezeichnung des Datenelements, in den zulässigen Werten oder in der Kodierung vor. Nun bestimmt der Gesetzgeber mit einer Novellierung des Personenstandsgesetzes in § 22 Absatz 3: „Kann das Kind weder dem weiblichen noch dem männlichen Geschlecht zugeordnet werden, so ist der Personenstandsfall ohne eine solche Angabe in das Geburtenregister einzutragen“. Keine der o.a. Definitionen bildet diese Rechtslage explizit ab; auch eine systematische Recherche nach anzupassenden Geschlechtsdefinitionen ist bei Fehlen einer entsprechenden Übersicht – z.B. in einem Metadata Repository – unmöglich.

3.1 Die strukturierte medizinische Dokumentation und „Value Sets“

Eine strukturierte Dokumentation unterstützt die Qualität der Patientenversorgung durch eine vollständige und eindeutige Erfassung von Daten. Als strukturiert wird hierbei eine Dokumentation von textuellen Inhalten über vordefinierte Datenelemente (z.B. „Geschlecht“) mit festgelegten Werteberei-

chen verstanden. Wertebereiche können über einen Datentyp (numerisch) oder eine Auswahlliste („männlich|weiblich|nicht festgelegt“) definiert sein. Datenelemente können zu komplexen Objekten zusammengefasst werden, z.B. zu Archetypes [Garde et al., 2006]. Auswahl und Festlegung von Datenelementen mit ihren Wertebereichen folgt dem Ziel des jeweiligen Dokumentationsverfahrens. Datenelemente mit Wertebereichen leiten sich von Fragestellungen und den zu ihrer Beantwortung erforderlichen Informationen ab [Löpprich et al., 2013]. International wird auch der Begriff „Value Set“ für eine Auswahlliste verwendet. Jede Ausprägung einer strukturierten Dokumentation – sei sie papiergestützt oder elektronisch – beinhaltet damit „Value Sets“. Dies umfasst Formulare im Krankenhaus wie einen Anforderungsschein für einen Herzultraschall ebenso wie die Case Report Forms der klinischen Forschung. Bei der Erstellung dieser Formulare ist jedes Mal neu die Festlegung von Datenelementen mit Wertebereichen sowie die Definition und die Bezeichnung der verwendeten Begriffe erforderlich.

Definitionen und Bezeichnungen von Datenelementen, Definitionen und Bezeichnungen von Werten aus Auswahllisten sowie Einträge für Kodes können aus Terminologien entnommen werden. Das trägt zur semantischen Interoperabilität einer strukturierten Dokumentation bei: Für die Kommunikation in der Versorgung (z.B. zwischen Krankenhäusern und Arztpraxen), für die mehrfache Nutzung von Daten in der Forschung (z.B. bei der Verknüpfung von Registern und klinischen Studien) oder für die Einbettung von Versorgung und Forschung in einen internationalen Rahmen (z.B. bei multinationalen Studien) [Ngouongo und Stausberg, 2011a]. Verknüpft eine Terminologie Begriffe mit Wertebereichen, kann dies komplett als Datenelement mit Value Set in ein Metadata Repository übernommen werden.

3.2 Klassifikationen versus Terminologien

Klassifikationen und Terminologien werden in unterschiedlicher Weise in ein Metadata Repository integriert [Ngouongo und Stausberg, 2011b]. Terminologien stellen Bezeichnungen und Definitionen von Datenelementen sowie von Einträgen in Auswahllisten, den Ausprägungen, bereit. Dies wären im Eingangsbeispiel die Bezeichnung „Geschlecht“ für das Datenelement sowie die Ausprägungen „männlich“, „weiblich“ und „unbekannt“. Zusätzlich zu den Ausprägungen definierte Kodes wie „1“ oder „m“ sind typischerweise nicht Bestandteil von Terminologien, sondern Konventionen im Rahmen der Spezifikation eines Dokumentationssystems. Denkbar wäre es, die Bezeichnungen der Ausprägungen selbst als Kode zu verwenden. In diesem Falle würden auch diese von der Terminologie bereitgestellt. Klassifikationen lassen sich aus Sicht eines Dokumentationssystems als komplexe Formen eines Wertebereichs auffassen. Für diesen liefern sie die zulässigen Ausprägungen sowie im Regelfall einen Kode. Bei einem Datenelement „Hauptdiagnose“ könnte die Be-

zeichnung „Hauptdiagnose“ aus einer Terminologie entnommen werden; der Wertebereich des Datenelements würde dann aus allen endständigen Klassen der Internationalen statistischen Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision, German Modification (ICD-10-GM) bestehen. In der Festlegung von Standards der medizinischen Dokumentation ergänzen sich somit Klassifikationen und Terminologien mit jeweils spezifischen Einsatzgebieten.

Problematisch ist die Integration von kontrollierten Vokabularen, wenn diese selbst ein komplexes Informationsmodell vorgeben. So lässt die ICD-10-GM in bestimmten Fällen eine Doppelkodierung von Diagnosen zu, kennt eine – optionale – Seitenangabe und wird im Kontext der Abrechnung nach § 295 SGB V um ein – obligates – Kennzeichen zur Diagnosesicherheit ergänzt. Entsprechend der Vorgaben der ICD-10-GM sind dann auch die Informationsmodelle in den verschiedenen Datenübermittlungsvereinbarungen formuliert; so besteht das Segment Nebendiagnose im Datenübermittlungsverfahren nach § 301 SGB V aus mindestens einer und maximal zwei Kombinationen eines Kodes der ICD-10-GM mit einer Seitenangabe [SGB-V, 2014]. Auch wenn sich die Integration von Klassifikationen und Terminologien in einem generischen Ansatz für ein Metadata Repository im Allgemeinen als unkritisch herauskristallisiert hat [Ngouongo et al., 2013], ist derzeit die Abbildung von Vokabular-spezifischen Informationsmodellen unklar. Hier besteht noch grundständiger Forschungsbedarf zur Erarbeitung einer generischen Lösung. Derzeit muss in Fällen wie der Doppelkodierung mit der ICD-10-GM mit Einzellösungen bei der Definition von Datenelementen in einem Metadata Repository gearbeitet werden.

3.3 Beispiel: SNOMED CT in einem Metadata Repository

Wie bereits eingeführt, bildet ein Metadata Repository die Plattform zur Definition von Datenelementen mit ihren Wertebereichen. Ein Terminologiesystem wie SNOMED CT stellt die einzelnen Bausteine (Begriffe, Bezeichnungen und Beziehungen) für diese Metadaten bereit. Damit wird eine Harmonisierung von Datenelementen und Wertebereichen über eine terminologische Fundierung gewährleistet [Ngouongo und Stausberg, 2013]. So kann aus SNOMED CT die Bezeichnung („Geschlecht eines Individuums“¹) und Definition (CONCEPTID 57312000) eines Datenelements, die Bezeichnung („männlich|weiblich|unbestimmtes Geschlecht“) und Definitionen („CONCEPTID 248153007| CONCEPTID 248152002|CONCEPTID 248154001“) der jeweiligen Werte des Value Sets sowie ggf. die Verknüpfung des Datenelements mit dem Value Set über die Relation „ist ein/e (Attribut)“ (CONCEPTID 116680003) entnommen werden. Auch Flexibilität in der Wahl der Bezeichnungen unterstützt das Metadata Repository, falls gewünscht. So könnte SNOMED CT für die Defini-

1 Die Beispiele aus SNOMED CT sind einer deutschen Übersetzung von 2004 entnommen.

tion der Metadaten genutzt werden; in den von Ärztinnen und Ärzten, Patientinnen und Patienten oder Bürgern genutzten Formularen würde eine kompatible Eingabeterminologie zum Einsatz kommen. Semantische Interoperabilität und Verständlichkeit wären so gleichermaßen gewährleistet.

3.4 Handlungsempfehlungen

In Deutschland wird eine Vielzahl unabgestimmter Festlegungen für Datenelemente eingesetzt. Mit dem Ziel einer Vereinheitlichung wird folgendes Vorgehen empfohlen:

1. Es sollte eine Ist-Erhebung zu denjenigen Festlegungen durchgeführt werden, die in Dokumentationsverfahren auf gesetzlicher Grundlage Anwendung finden. Dies schließt Versorgung und Forschung gleichermaßen ein. Die Ist-Erhebung sollte unter Verwendung eines Metadata Repository erfolgen, um darüber die Ergebnisse für eine systematische Bewertung, für eine Harmonisierung sowie für eine zielgerichtete Ergänzung verfügbar zu machen.
2. Als Voraussetzung und Mittel zu dieser Pflege von Datenelementen wird die Implementierung eines versorgungs- und forschungsübergreifenden nationalen Metadata Repositories empfohlen. Dies kann auf Basis und ggf. auch unter Ausbau des Metadata Repository der TMF erfolgen [TMF, 2009–2012]. Übergangsweise ist auch eine abgestimmte Parallelität von zwei nationalen Metadata Repositories denkbar, von denen eines mit dem Schwerpunkt Forschung und ein weiteres mit dem Schwerpunkt Versorgung betrieben wird.
3. Für die Forschung lässt sich in der TMF eine geeignete Institution zur Koordinierung und auch zur Verantwortung entsprechender Aktivitäten identifizieren; diese Institution fehlt in der Versorgung. Hierfür lassen sich gleichwertig eine Erweiterung der Aufgaben der TMF bei zusätzlicher Finanzierung als auch der Neuaufbau eines entsprechenden Bereichs diskutieren. Letzterer sollte an ein Hochschulinstitut der Medizinischen Informatik angebunden sein, um eine der Innovativität angemessene Verzahnung von grundständiger Forschung zu Metadaten mit dem Management von Metadaten sicherzustellen.
4. Zur Verknüpfung von Metadaten und kontrollierten Vokabularen gibt es vielversprechende Ansätze; eine vollständige Durchdringung dieser Verknüpfung steht allerdings noch aus. Zur Deckung des Forschungsbedarfs sollte eine projektbezogene Förderung über die nächsten zehn Jahre initiiert werden. Hierfür könnten etablierte Förderprogramme des Bundesministeriums für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG) ausgebaut werden, falls eine zielgerichtete Koordinierung erfolgt. Alternativ ist auch an ein Schwerpunktprogramm der DFG zu denken.

5. Angesichts des enormen Aufwands zur Entwicklung einer umfassenden Terminologie für das Gesundheitswesen ist SNOMED CT – trotz methodischer Vorbehalte und mäßigem Erfolgs – als kontrolliertes Vokabular konkurrenzlos. SNOMED CT sollte daher in einer deutschen Fassung für die gemeinfreie Verwendung in einem Metadata Repository und damit für alle Belange von Versorgung und Forschung verfügbar gemacht werden. Dies ist nicht mit einer Abbildung aller gesundheitsbezogenen Sachverhalte mit SNOMED CT gleichzusetzen, sondern sollte in einer Übernahme ausgewählter „Value Lists“ oder einzelner Begriffe und ihrer Bezeichnungen aus SNOMED CT liegen, dort, wo dies sinnvoll erscheint.

Referenzen

- [Bundesanzeiger, 2014] Bundesministerium für Gesundheit (2014): Bekanntmachung Aktualisierter einheitlicher onkologischer Basisdatensatz der Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT) und der Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (GEKID).
- [Bürkle, T., 2000] Bürkle, T. (2000): Can we classify medical data dictionaries? *Stud Health Technol Inform* 77: S. 691–695.
- [DIVI, 2010] Deutsche interdisziplinäre Vereinigung für Intensiv- und Notfallmedizin (2010): Kerndatensatz Notaufnahme der DIVI V1.0.
- [Garde, S. et al., 2006] Garde, S. et al. (2006): Ubiquitous information for ubiquitous computing: expressing clinical data sets with openEHR archetypes. *Stud Health Technol Inform* 124: S. 215–220.
- [ISO/IEC, 2013] International Organization for Standardization (2013): Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes.
- [KBV, 2014] KBV. Schlüsselstabellen. <http://applications.kbv.de/keytabs/ita/schluesseltabellen.asp> Letzter Zugang: 2014-08-11.
- [KHEntgG, 2013] KHEntgG. Anlage zur Vereinbarung über die Übermittlung von Daten nach § 21 Abs. 4 und Abs. 5 KHEntgG – Daten nach § 21 KHEntgG – Version 2014 für das Datenjahr 2013. Fortschreibung vom 26. November 2013. Im Internet: http://www.dkgev.de/media/file/15428.RS407-13_Daten-%C2%A721-KHEntgG_Version-2014_A1.pdf Letzter Zugang: 2014-08-11.
- [Kupermann, G.J. et al., 1991] Kupermann, G.J. et al. (1991): HELP: A Dynamic Hospital Information System. (1.Ed.). Springer-Verlag.
- [Leiner, F. und Haux, R., 1996] Leiner, F. und Haux, R. (1996): Systematic planning of clinical documentation. *Methods Inf Med* 35 (1): S. 25–34.
- [Linnarsson, R. und Wigertz, O., 1989] Linnarsson, R. und Wigertz, O. (1989): The data dictionary – a controlled vocabulary for integrating clinical databases and medical knowledge bases. *Methods Inf Med* 28 (2): S. 78–85.
- [Löpprich, M. et al., 2013] Löpprich, M. et al. (2013): Inhalt und Aufbau eines Registerprotokolls. Vortrag im Rahmen von: 58. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS), Lübeck.
- [METeOR, 2014] METeOR. METeOR – Metadata Online Repository. <http://meteor.aihw.gov.au> Letzter Zugang: 2014-08-11.
- [NCI, 2010] NCI. caDSR Wiki. <https://wiki.nci.nih.gov/display/caDSR/caDSR+Wiki> Letzter Zugang: 2014-08-11.
- [Nguouongo, S.M. et al., 2013] Nguouongo, S.M. et al. (2013): The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research? *J Biomed Inform* 46 (2): S. 318–327.
- [Nguouongo, S.M. und Stausberg, J., 2011a] Nguouongo, S.M. und Stausberg, J. (2011a): Harmonisierung von Datenbeständen der Medizin | Ein Weg zu besserer Qualität und höherer Effizienz empirischer Forschung *Forum der Medizin-Dokumentation und Medizin-Informatik* 3/2011: S. 129–132.

- [Nguouongo, S.M. und Stausberg, J., 2011b] Nguouongo, S.M. und Stausberg, J. (2011b): Integration of classifications and terminologies in Metadata registries based on ISO/IEC 11179. *Stud Health Technol Inform* 169: S. 744–748.
- [Nguouongo, S.M. und Stausberg, J., 2013] Nguouongo, S.M. und Stausberg, J. (2013): Terminology-based documentation systems: a systemantic comparison of four different approaches. *Stud Health Technol Inform* 190: S. 97–99.
- [SGB-V, 2014] SGB-V. Datenübermittlung nach § 301 Abs. 3 SGB V, ergänzt um die Festlegungen der Vereinbarung nach § 120 Abs. 3 SGB V. Stand: Januar 2014. Im Internet: http://www.dkgev.de/media/file/15929_RS060-14_Anlage-SGBV_v301_Dokumentation.pdf Letzter Zugang: 2014–08–11.
- [SQG, 2014] SQG. Sektorenübergreifende Qualität im Gesundheitswesen. <https://www.sgg.de> Letzter Zugang: 2014–08–11.
- [Stausberg, J. und Hucklenbroich, P., 1994] Stausberg, J. und Hucklenbroich, P. (1994): Eine Struktur fallbezogener medizinischer Informationen als Grundlage wissensbasierter Systeme. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 25: S. 140–148.
- [Stausberg, J. et al., 2009] Stausberg, J. et al. (2009): Foundations of a metadata repository for databases of registers and trials. *Stud Health Technol Inform* 150: S. 409–413.
- [Stausberg, J. et al., 1995] Stausberg, J. et al. (1995): Terminological reference of a knowledge-based system: the data dictionary. *Medinfo* 8 Pt 1: S. 157–161.
- [TMF, 2009–2012] TMF. D021–01 Metadata Repository. http://www.tmf-ev.de/Themen/Projekte/D021_01_Metadata_Repository.aspx Letzter Zugang: 2014–08–11.