

Ausblick semantische Standards für eHealth in der Schweiz

V 2.0

17.11. 2015

Impressum

Autoren: Juerg Peter Bleuer, Healthvidence GmbH
Hans Rudolf Straub, Semfinder AG

Stand: 17. November 2015: V 2.0

Korrespondenz:

*Dr. med. Juerg P. Bleuer, MPH
Healthvidence GmbH
PF 6551
3001 Bern
E-Mail: bleuer@healthvidence.ch*

*Med. pract. Hans Rudolf Straub,
Semfinder AG
Hauptstrasse 53
8280 Kreuzlingen
E-Mail: straub@semfinder.com*

Inhalt

1	Management Summary	7
2	Einleitung	8
2.1	Ausgangslage und Motivation	8
2.2	Zweck des Dokuments	8
2.3	Angesprochene Leserschaft	8
3	Zu diesem Dokument	9
3.1	Abgrenzung	9
3.2	Relevante Publikationen von eHealth Suisse	9
3.3	Übersetzungen	9
3.4	Aufbau dieses Dokumentes	9
4	Anwendungsfälle	11
4.1	Vorbemerkung	11
4.2	Storyboard	11
4.3	Diskussion	13
4.3.1	Model of Information und Model of Meaning	13
4.3.2	Klassierung und Granularität	14
4.3.3	Codes.....	15
4.3.4	Translations sind unvermeidbar.....	15
5	Semantische Grundlagen	16
5.1	Uneinheitliche Bezeichnungen	16
5.2	Semiotisches Dreieck	16
5.2.1	Eigenschaften der Begriffe, die bei Wörtern nicht vorkommen	18
5.2.2	Kommunikation	19
5.3	Ordnungssysteme, Klassifikationen, Codesysteme, Terminologien usw.	19
5.3.1	Vorbemerkung	19
5.3.2	Ordnungssysteme ohne explizite innere Struktur	19
5.3.3	Hierarchische Ordnungssysteme.....	20
5.3.4	Semantische Achse oder Dimension, semantischer Freiheitsgrad	20
5.3.5	Unidirektionalität, Multihierarchie	21
5.3.6	Mehrdimensionale Ordnungssysteme	22
5.3.7	Multifokale Ordnungssysteme.....	23
5.3.8	Ontologien.....	24
5.4	Semantische Analyse von Diagnoseinformationen	25
5.4.1	Beispiel Abdominalhernien ICD-10 WHO	25
5.4.2	Beispiel Abdominalhernien ICD-10 GM	26
5.4.3	Multifokale Darstellung der semantischen Achsen	28

5.4.4	Abbildung von Inhalten auf die verschiedenen Typen von Ordnungssystemen.....	28
5.5	Probleme bei der Abbildung einer komplexen Semantik in ein hierarchisches System	29
5.5.1	Nicht darstellbare Information	29
5.5.2	Kombinatorische Explosion.....	30
5.5.3	Verteilung einer singulären Information auf mehrere Codes	30
5.6	Probleme bei der Abbildung einer komplexen Semantik in ein multidimensionales System	30
5.7	Mapping	31
5.7.1	Direktes Mapping	31
5.7.2	Mapping über eine Referenz.....	32
5.7.3	Mapping über geführte Eingabe bzw. NLP (Mehrfachcodierung).....	33
5.7.4	Mapping (Zusammenfassung)	34
6	Data Entry: Was sind die Anforderungen bezüglich Code-Eingabe beim Anwender? Unterschied zwischen technischen und klinischen Daten	35
6.1	Herausforderungen für das Data Entry:	35
6.2	Lösungsvarianten für das Data Entry:.....	36
7	Interoperabilitäts-Niveaus	37
7.1	Technische Interoperabilität.....	37
7.2	Syntaktische Interoperabilität	37
7.3	Semantische Interoperabilität	37
7.4	Prozessinteroperabilität	38
7.5	Verschiedene Modelle der Interoperabilität	38
7.5.1	Modell der EHR Interoperability Work Group.	38
7.5.2	Levels of Conceptual Interoperability Model (LCIM).....	38
8	Expertengespräche	40
8.1	Einleitung.....	40
8.2	Methodik	40
8.2.1	Teilnehmer der Expertengespräche	40
8.2.2	Fragenkatalog für die Expertengespräche.....	41
8.2.3	Diskussion und Rapportierung an die Expertengruppe Semantik eHealth Suisse.....	41
8.3	Resultate.....	41
8.3.1	Zielsetzung.....	41
8.3.2	Adressaten der Umfrage / Vorgehen	41
8.3.3	Fragenkatalog	42
8.3.4	Veränderte Erwartungen an das vorliegende Dokument.....	42
9	Sicht der Autoren	43
9.1	Ziele.....	43
9.2	Fachlich semantische Ebene	44

9.3	Umfeld.....	45
10	Empfehlungen / Massnahmen für die Umsetzung	46
10.1	Ausgangspunkt.....	46
10.2	Eigenschaften von Implementierungsleitfäden	46
10.3	Pflege von Austauschformaten und Referenzimplementation	47
10.4	Puffer	47
10.5	Umfrage	48
10.6	Schlussfolgerung.....	48
11	Anhang	49
11.1	Fragenkatalog (Vorschlag der Autoren ergänzt mit Vorschlägen aus dem ersten Expertengespräch und aus der Expertengruppe Semantik)	49
11.1.1	Einleitung	49
11.1.2	Fragen zur aktuellen Nutzung von strukturiertem und codiertem Informationsaustausch	49
11.1.3	Fragen zu den Erwartungen an einen zukünftige strukturierten und codierten Informationsaustausch.....	50
11.1.4	Fragen zu den Anforderungen der Befragten an eine strukturierte und codierte Informationsverarbeitung.	50
11.1.5	Weitere Fragen	51
11.2	Adressaten für eine Umfrage (Vorschlag der Autoren ergänzt mit Vorschlägen aus dem ersten Expertengespräch und aus der Expertengruppe Semantik) ..	51
11.3	Anmerkung zur Umfrage	54
11.4	Glossar	54
11.4.1	Begriff.....	54
11.4.2	Bezeichnung	54
11.4.3	Binding	54
11.4.4	Code.....	54
11.4.5	Code System.....	55
11.4.6	Granularität	55
11.4.7	Klassifikation	55
11.4.8	Mapping	56
11.4.9	Ontologie.....	56
11.4.10	Ordnungssystem	57
11.4.11	Präkoordination / Postkoordination	57
11.4.12	Prozessinteroperabilität	57
11.4.13	Referenzterminologie	57
11.4.14	Semantik	57
11.4.15	Semantische Interoperabilität	58
11.4.16	Semiotisches Dreieck	58

11.4.17	Terminologie	58
11.4.18	Value Set.....	59
11.5	Synopsis der semantischen Bezeichnungen und Begriffe	61
11.6	Synopsis der Ordnungssysteme.....	63

1 Management Summary

Mit beigezogenen Experten sowie Vertretern der Expertengruppe Semantik eHealth Suisse wurde eine Liste von Adressaten sowie ein Fragenkatalog für eine Erhebung zur heutigen und zukünftigen Nutzung von semantischen Standards in der Schweiz erarbeitet.

In den Expertengesprächen wurde das Fehlen einer Semantikstrategie als eines der Grundprobleme lokalisiert. Da Terminologie und Datenmodelle nicht unabhängig voneinander betrachtet werden dürfen, sprechen wir anstelle von „Semantikstrategie“ von „Interoperabilitätsstrategie“. Diese muss explizit das Ziel der Prozessinteroperabilität verfolgen.

Aus der Forderung nach Prozessinteroperabilität geht hervor, dass die ganze Kette „Mensch (Autor) – Maschine, Maschine – Maschine und Maschine – Mensch (Empfänger)“ betrachtet werden muss, also auch Dateneingabe und Anzeige beim Empfänger.

Implementierungsleitfäden für Austauschformate werden auch zukünftig von unterschiedlichen Autoren erstellt werden und in unterschiedlichen Knowledge Domains beheimatet sein. Damit diese Implementationsleitfäden zu Austauschformaten die Prozessinteroperabilität fachgebietsübergreifend sicherstellen, empfehlen wir entsprechende Guidelines; als Herausgeber sehen wir das Koordinationsorgan eHealth Suisse.

Erstellte Austauschformate benötigen Pflege; dafür müssen die benötigten Ressourcen bereit gestellt und die Verantwortlichkeiten bestimmt werden; die Federführung sehen wir z.B. bei eHealth Suisse.

Sachzwänge werden es immer wieder notwendig machen, bei den verwendeten Ordnungssystemen in Teilbereichen Anpassungen vorzunehmen und/oder zumindest mit zeitlichen Verschiebungen bei der Umsetzung leben zu müssen. Dies trifft sowohl für einzelne Domains unter einander als auch für die Beziehung zwischen eHealth in der Schweiz und im Ausland zu.

Daraus ergibt sich der Bedarf nach „Puffer“ für das Management von Abweichungen von Policies, Datenmodellen und verwendeten Begriffen und deren Codierung.

Als technisches Mittel bietet sich eine Weiterentwicklung eines Terminologie Servers mit zusätzlichen Funktionalitäten an, welche über die Verwaltung von Relationen zwischen Referenzterminologien¹, Value Sets und Bindings sowie die Pflege der einzelnen Ordnungssysteme hinausgeht: Ein solches System existiert nach unserem Wissen weltweit nicht. Die Autoren empfehlen, eine entsprechende Konzeptarbeit zu veranlassen. Für den Betrieb müsste landesweit eine einzige Stelle verantwortlich sein: diese würde somit auch die Inhalte der Referenzterminologien für das schweizerische Gesundheitswesen definieren. Eine föderalistische Alternative sehen wir nicht.

Für den Laborbereich sind aktuell (2015) in der Schweiz LOINC und SNOMED CT daran, sich als Referenzterminologien zu etablieren; es gibt keinen Grund, dies in Frage zu stellen; SNOMED CT gilt auch für weitere Anwendungen als erste Wahl. Grundsätzlich unterstützen wir dies; allerdings sind noch Fragen betreffend Data Entry bei SNOMED CT zu studieren.

¹ Genau genommen müsste es „Referenz-Ordnungssystem“ heissen (s. a. Eintrag im Glossar). Diese Bezeichnung ist allerdings unüblich; wir verwenden deshalb auch „Referenzterminologie“.

2 Einleitung

2.1 Ausgangslage und Motivation

Der Bundesrat hat am 27. Juni 2007 die Strategie „eHealth“ Schweiz genehmigt. Deren erfolgreiche Umsetzung gelingt nur dann, wenn die Prozessinteroperabilität der beteiligten Institutionen gegeben ist. Diese wiederum erfordert einen strukturierten Datenaustausch mit semantischer Interoperabilität der beteiligten IT-Systeme: Nur so kann Information weitgehend automatisch interpretiert und damit Medienbruch-frei ohne „händische“ Intervention institutionsübergreifend verarbeitet werden.

Das Ziel, das mit der "semantischen Interoperabilität" verfolgt wird, ist eine Kommunikation mit eindeutiger Zuordnung zwischen allen verwendeten Bezeichnungen und den damit gemeinten Sachverhalten.

2.2 Zweck des Dokuments

Das vorliegende Strategiepapier diskutiert semantische Fragen und vermittelt die dafür notwendigen Grundlagen. Die enthaltenen Empfehlungen basieren auf internationalen Standards und Profiles sowie Expertenempfehlungen.

2.3 Angesprochene Leserschaft

Das vorliegende Dokument richtet sich an alle Fachleute im Gesundheitswesen und in der Gesundheitspolitik. Angesprochen sind insbesondere auch Nicht-Informatiker.

Aus Gründen der Lesbarkeit wird bei Personenbezeichnungen darauf verzichtet, sowohl die weibliche als auch die männliche Form anzuführen.

3 Zu diesem Dokument

3.1 Abgrenzung

Folgende Inhalte sind auftragsentsprechend im vorliegenden Dokument NICHT enthalten:

- Konkrete Empfehlung für bestimmte Standards
- Festlegungen von Wertebereichen (Value-Sets)
- Resultate einer allfälligen Schweiz-weiten Umfrage; möglicherweise wird eine solche Umfrage zu einem späteren Zeitpunkt erfolgen.

3.2 Relevante Publikationen von eHealth Suisse

Verabschiedete Dokumente:

<http://www.e-health-suisse.ch/umsetzung/00146/00148/00238/index.html?lang=de>

- Evaluationskriterien für Semantische Standards
- Empfehlungen I Semantik und Metadaten
- Die Methodik zur Evaluation von semantischen Standards bzw. die entsprechenden Dokumente von „eHealth Suisse“.
- Ist-Analyse Semantik 2011
- Inputpapier Semantik und Metadaten in Englisch, 2011

3.3 Übersetzungen

Diese Version des Dokuments ist ausschliesslich in Deutsch verfügbar.

3.4 Aufbau dieses Dokumentes

Die Diskussion semantischer Aspekte ist abstrakt; aus diesem Grund werden zu Beginn verschiedene Aspekte der semantischen Interoperabilität anhand von Anwendungsfällen illustriert und die damit verbundenen Herausforderungen und Probleme lokalisiert. Ein darauf folgendes Grundlagenkapitel vermittelt das Verständnis semantischer Fragen.

Semantische Fragen betreffen nicht nur die Terminologie sondern auch die Informationsstruktur; das Grundlagenkapitel geht auch auf diesen Aspekt ein. Die verwendeten Fachausdrücke werden in einem Glossar im Anhang aufgeführt. Eine Tabelle zeigt andernorts verwendete Bezeichnungen in einer Übersicht.

Basierend auf Experteninterviews wird anschliessend an das Grundlagenkapitel ein mögliches Vorgehen zur Erhebung des heutigen Einsatzes und der Nutzung von semantischen Standards in der Schweiz entwickelt.

Am Schluss begründen die Autoren ihre Sichtweise und sie geben ihre Empfehlungen ab.

Im Interesse der leichten Lesbarkeit wird bewusst auf einen wissenschaftlichen Schreibstil verzichtet. Quellenangaben werden in Fussnoten angegeben; für den Leser ist dies angenehmer als ein Refe-

renzverzeichnis am Schluss des Dokumentes. Die Autoren sprechen den Leser in der „wir“-Form an, insbesondere bei Stellungnahmen und Wertungen.

4 Anwendungsfälle

4.1 Vorbemerkung

Das nachfolgende Storyboard geht von dem in CDA-CH II beschriebenen Autounfall aus; allerdings projizieren wir den Fall in eine Zukunft, in der sowohl des elektronische Patientendossier als auch die strukturierte Kommunikation zwischen allen beteiligten Akteuren Realität ist. Aus dem ursprünglichen Fall ist einiges weggelassen, anderes ist zwecks Illustration der semantischen Aspekte neu hinzugefügt.

Das beschriebene Geschehen nimmt einen glücklichen Verlauf; eine umfangreiche pflegerische Nachbetreuung ist nicht nötig und die Pflege wird deshalb nicht explizit erwähnt. Es wäre allerdings falsch daraus zu schliessen, semantische Fragen wären nicht von Belang für die Pflege und weitere im Beispiel ebenfalls nicht vorkommende Bereiche der Gesundheitsversorgung.

Semantik und Struktur bilden die Grundpfeiler für jede Kommunikation (im Gesundheitswesen); im Interesse des einfacheren Verständnisses beschränken wir uns auf ein überschaubares Szenario.

4.2 Storyboard

Herrn Max Muster befällt während einer Autofahrt mit einem Bekannten plötzlich Schwindel, Unwohlsein und Herzrasen. Er bremst den Wagen ab, bevor er jedoch seitlich auf das Trottoir fahren kann, prallt der hinter ihm fahrende Wagen in sein Heck. Der Auffahrunfall ist zwar nicht heftig und auch das Schwindelgefühl lässt bereits nach. Weil Herr Muster aber doch einen ziehenden Schmerz im Nacken verspürt, besteht sein Begleiter darauf, dass er sich in der Notfallstation des nahegelegenen kantonalen Spitals untersuchen lässt.

Der Notfallarzt des Kantonalen Spitals, Dr. E. Mergency, lässt sich von Max Muster den Hergang des Unfalls beschreiben. Er veranlasst eine Röntgenkontrolle der Halswirbelsäule: Die Röntgenverordnung erstellt er in seinem KIS, welches die Verordnung elektronisch an die Radiologie im Haus übermittelt (1). Die Bilder und der Befundbericht werden aus der Radiologie elektronisch an das KIS geschickt (2). Der Befundbericht ist strukturiert; die einzelnen Befunde und die Beurteilung sind somit in strukturierter Form vorhanden. Die Untersuchung zeigt eine leichte rechtskonvexe Skoliose und glücklicherweise keine Hinweise auf Verletzungen. Routinemässig wird auch eine Laboruntersuchung des Blutes durchgeführt; auch hier erfolgen Auftrag an das Labor und Rückmeldung der Resultate elektronisch und strukturiert (3,4).

Aufgrund der klinischen und radiologischen Befunde stellt Dr. E. Mergency die vorläufige Diagnose einer Distorsion der Halswirbelsäule Grad II der QTF-Klassifikation. Er behandelt die mittelstarken Nackenschmerzen mit einem einfachen und gut verträglichen Schmerzmittel und einem Medikament zur Muskelentspannung (Paracetamol 1000 mg 3 mal tgl. ; Tizanidin 4 mg 3 mal tgl.). Im KIS wird ein elektronisches Rezept erstellt, womit die verordneten Medikamente in strukturierter Form an die Apotheke übermittelt werden kann (5). Dies ermöglicht dort die Weiterverarbeitung ohne Medienbruch.

Herr Muster erwähnt, dass er wegen zu hohem Blutdruck seit Jahren ein Medikament einnimmt, kann sich aber nicht an den Namen erinnern. In dem elektronischen Patientendossier von Herrn Muster sieht Dr. E. Mergency, dass der Patient mit dem blutdrucksenkenden Mittel Lisinopril HCT 10/12.5 mg täglich behandelt wird. Zudem geht hervor, dass eine Hyperthyreose, also eine Überfunktion der Schilddrüse bekannt ist, welche bis vor 6 Monaten mit dem Medikamentenwirkstoff Carbimazol 5mg einmal täglich behandelt wurde. Herr Muster hat dem Notfallarzt davon nichts erzählt, da er keinen

Zusammenhang mit der aktuellen Situation erkennt und deshalb eine Erwähnung nicht für nötig hält. Dr. Mergency vermutet, dass der Schwindel und das Unwohlsein, welche zu der Auffahrkollision führten, durch eine Störung des Herzrhythmus verursacht wurden, wie sie gelegentlich bei Patienten mit Schilddrüsenüberfunktion vorkommt. Obwohl der Arzt bei der körperlichen Untersuchung dafür keine Anhaltspunkte findet, lässt er ein EKG durchführen, welches allerdings einen Normalbefund zeigt. Er empfiehlt Herrn Muster, sich bei seinem Hausarzt, Dr. A. Bereit, zu melden. Diesem schickt er elektronisch einen strukturierten ambulanten Austrittsbericht (6) in welchem er seinen Verdacht beschreibt. Ebenfalls elektronisch übermittelt er das EKG. Schliesslich empfiehlt er weitere Abklärungen.

Zwei Tage nach dem Auffahrunfall meldet sich Herr Muster bei seinem Hausarzt Dr. A. Bereit. Die Schmerzen im Nacken bestehen noch immer; sie haben eher an Stärke zugenommen, lassen sich aber mit der vom Spital verordneten Schmerzmedikation einigermaßen beherrschen. Dr. Bereit führt eine Blutentnahme zur Kontrolle der Schilddrüsenhormone im Serum des Patienten durch. Die Bestimmung erfolgt in einem externen Labor; der Datenaustausch zwischen Labor und Hausarzt erfolgt sowohl für den Laborauftrag als auch bei der Übermittlung der Resultate elektronisch und strukturiert (7,8).

Angesichts der vorangegangenen Schilddrüsenüberfunktion und dem anfallsweise aufgetretenen Herzrasen zieht Dr. Bereit ein paroxysmales Vorhofflimmern in Betracht und weist Herrn Muster zur weiteren Abklärung an den Kardiologen Dr. A. Orta; auch hier erfolgt die Zuweisung elektronisch und strukturiert (9). Vorsichtshalber, um eine Hirnembolie zu vermeiden, leitet er eine „Blutverdünnung“ mit einem Acetylsalicylat-Präparat ein. Wiederum wird elektronisch, diesmal in der Praxis-Software, wird ein elektronisches und strukturiertes Rezept erstellt (10). Schliesslich überträgt Dr. Bereit die Untersuchungsbefunde des kantonalen Spitales inklusive der Beurteilung von EKG und Röntgenuntersuchung der Halswirbelsäule sowie die Medikamenten-Verordnungen in Anwesenheit des Patienten direkt aus der Praxis-Software in das elektronische Patientendossier des Patienten (11).

Der Kardiologe Dr. A. Orta möchte sich über die ganze Vorgeschichte des Patienten orientieren; Max Muster erlaubt ihm deshalb den Zugriff auf sein elektronisches Patientendossier. Damit kann Dr. A. Orta u.a. das im Kant. Spital durchgeführte EKG einzusehen. Ebenfalls verfügbar sind Anamnese, frühere Therapien etc.; alles in strukturierter Form (12).

Zum Nachweis sporadisch auftretender Herzrhythmusstörungen führt Dr. Orta daraufhin eine Langzeit-EKG Untersuchung (sog. R-Test) durch. Tatsächlich finden sich dabei Hinweise auf intermittierende Episoden mit raschem und arrhythmischem Puls. Dr. Orta stellt die Diagnose „paroxysmales tachykardes Vorhofflimmern“ und verordnet Marcoumar-Tabletten zur „Blutverdünnung“. Auch jetzt erfolgt die Rezepterstellung elektronisch und strukturiert (13).

Die Befunde, die Beurteilung und die Verordnung übermittelt Dr. Orta elektronisch und strukturiert an Dr. Bereit (14).

Max Muster begibt sich nach einigen Tagen wieder zu seinem Hausarzt Dr. Bereit. Die Laborwerte haben inzwischen den Verdacht auf eine erneute Schilddrüsen-Überfunktion bestätigt. Diese verursacht mit grösster Wahrscheinlichkeit die Herzbeschwerden des Patienten. Dr. Bereit orientiert seinen Patienten über die nun vorliegenden Befunde aus der Konsultation beim Kardiologen. Die „Blutverdünnung“ muss weitergeführt werden, bis sich der Herzrhythmus sicher stabilisiert hat und Herr Muster wird auch mit einem Medikament zur Behandlung der Schilddrüsen-Überfunktion behandelt (Carbimazol-Tabletten). Die Medikation mit Dafalgan und Sirdalud Tabletten wird weitergeführt. Diese Verordnungen werden wiederum strukturiert in das elektronische Patientendossier (15) übertragen. Ein neues Rezept wird erstellt, welches strukturiert die aktuell verordneten Medikamente enthält (16).

Das Happyend: Unter der medikamentösen Behandlung mit Carbimazol normalisieren sich die Schilddrüsen-Werte allmählich und Herr Muster hat auch keine Schwindelanfälle oder Episoden von Herzklopfen oder Herzrasen mehr. Eine erneute kardiologische Kontrolle nach 2 Monaten ergibt im R-Test keine weiteren Rhythmusstörungen mehr. Auch kann in der transoesophagealen Echografie keine Thrombusbildung nachgewiesen werden. Damit kann nun auch die medikamentöse Blutverdünnung gestoppt werden. Auch die Nackenschmerzen haben sich nun stark gebessert, eine weitere Therapie ist zum Glück nicht erforderlich.

4.3 Diskussion

4.3.1 Model of Information und Model of Meaning

Im vorangehenden Kapitel haben wir bereits 16 Kommunikationsprozesse lokalisiert; deren Inhalt wird – wie im eHealth Kontext üblich - als „**strukturiert**“ bezeichnet.

Obschon wir uns im vorliegenden Dokument vor allem mit der Semantik beschäftigen, sei zur Strukturierung folgendes erwähnt:

Zu unterscheiden ist zwischen der Strukturierung der Eingabe und der Strukturierung der Nachricht: Eine strukturierte Nachricht setzt nicht zwingend eine strukturierte Eingabe voraus: Mit Natural Language Processing (NLP) Methoden ist es prinzipiell möglich, aus Freitext oder sogar gesprochenem Text ein strukturiertes Dokument zu erstellen. Noch sind solche Methoden unvollkommen; Kapitel 5.7.3 und 6.2 gehen auf NLP näher ein.

Bis auf weiteres werden medizinische Inhalte wohl nicht in jedem Fall vollständig strukturiert abgebildet werden können. Bestimmte Sachverhalte sind zu komplex; man denke beispielsweise an eine psychiatrische Anamnese.

Strukturierung der Information ist zwar eine notwendige, jedoch nicht hinreichende Voraussetzung für eine Realisierung der angestrebten Medienbruch-freien institutionsübergreifenden Informationsverarbeitung ohne „händische“ Intervention: Dies wird anhand eines einfachen Beispiels klar: Ein medizinischer Laie kann in einem Überweisungsschreiben zwar eine Gliederung in Anamnese, Befunde, Verdachtsdiagnosen, Fragestellung etc. erkennen; in Ermangelung der Kenntnis der medizinischen Terminologie versteht er den Inhalt des Schreibens wahrscheinlich gleichwohl nicht.

Eine – vom Autor und Adressat eines Dokumentes – gleich verstandene Terminologie ist daher ebenfalls eine notwendige Voraussetzung. Ein Beispiel möge auch dies illustrieren: Oft versteht man bei einer fremdsprachigen Konversation die relevanten Begriffe (z. B. weil es englische Fremdwörter sind); wegen mangelnden Sprachkenntnissen bleibt jedoch der Zusammenhang und damit die Bedeutung der Aussagen unklar.

Rector et al. definieren Electronic Health Records als Datenstrukturen, welche bestimmten Informationsmodellen folgen: Zusätzlich werden „Terminology or concept models“ benötigt, welche die „meaning of what is stored“ repräsentieren.

Ein solches „model of meaning“ stellt eine Repräsentation unseres Verständnisses der realen Welt dar und ermöglicht damit die Argumentation; Rector et. al. stellen fest: „... the model of meaning ... is a model of our conceptualisation of the world – of patients, their illnesses².“

Sowohl bei der Entwicklung von HL7 V 2 und V3 und damit auch CDA wurde das Informationsmodell und damit die Struktur unabhängig von irgendwelchen Terminologien entwickelt. SNOMED CT wiederum wurde explizit Syntax-neutral erstellt, damit sie mit jedem Informationsmodell und somit mit beliebigen Datenstrukturen verwendet werden könne. Gerechtfertigt wird die vollständige Trennung von Syntax- und Terminologieentwicklung mit der Notwendigkeit der Verwendung mit zahlreichen Legacy-Systemen, mit unterschiedlichen Sprachen und unterschiedlichen nationalen Anforderungen³.

² Rector, AL, Qamar R, Marley T. Binding Ontologies and Coding Systems to Electronic Health Records and Messages. Proceedings of KR-MED-2006; Baltimore, USA.

³ Da wir hier vom trilateralen Modell des semiotischen Dreiecks ausgehen (s. Kapitel 5.2), müsste anstelle von „Syntax“ eigentlich von „Syntaktik“ gesprochen werden. Der Begriff der Syntaktik stammt von Charles W. Morris: Syntaktik befasst sich mit der Beziehung von Zeichen zu anderen Zeichen. Allerdings wird die Unterscheidung zwischen „Syntax“ und „Syntaktik“ häufig unterlassen. Wir kommen nicht umhin „Syntax“ und „Syntaktik“ synonym zu verwenden; weil in den zitierten Quellen beide Be-

Struktur und Terminologie sind jedoch voneinander abhängig; auch bei jeder gesprochenen Sprache haben sich Syntax und Terminologie gemeinsam entwickelt. Dies trifft auch für das Zusammenwirken von Terminologien und formalen Sprachen zu.

Ein einfaches Beispiel illustriert dies: Nehmen wir an, das Informationsmodell eines elektronischen Patientendossier sehe vor, das bestimmte Sachverhalte durch geeigneten Codes (mit einer definierten Bedeutung) zu repräsentieren und die Abwesenheit eines bestimmten Sachverhaltes durch den entsprechenden Code, verbunden mit einem Marker für die Negation darzustellen.

ICD-10-GM hat einen Code O73.0 mit der Bedeutung „Retention der Plazenta ohne Blutung“ Die Negation der Blutung ist mit „ohne Blutung“ bereits Teil der Terminologie; dies passt nicht zum obigen Informationsmodell.

Das Beispiel zeigt deutlich, dass Terminologien und Codes (s. Kapitel 4.3.3) nicht unabhängig von der Datenstruktur diskutiert werden können. Umgekehrt können aber auch die Datenstrukturen – wie das oben angeführte Beispiel zeigt – nicht unabhängig von den Terminologien und Codes behandelt werden.

4.3.2 Klassierung und Granularität

In allen Dokumenten wird es einen Eintrag „Max Muster“ geben. Es ist selbstverständlich, dass mit der Bezeichnung „Max Muster“ der Patient Max Muster als Individuum gemeint ist. Demgegenüber bezeichnet „paroxysmales tachykardes Vorhofflimmern“ nicht den Einzelfall des Vorhofflimmerns bei Herrn Muster, sondern die Klasse aller derartiger Vorhofflimmern. Das „paroxysmale tachykarde Vorhofflimmern“ von Herrn Muster ist eine Instanz der Klasse „paroxysmales tachykardes Vorhofflimmern“. Diese ist eine Unterklasse aller Vorhofflimmern, diese eine Unterklasse aller Herzrhythmusstörungen und diese wiederum eine Unterklasse aller Herzprobleme.

Folgende Parallelität ist festzuhalten: In der objektorientierten Programmierung (OOP) spricht man auch von Objekttypen und davon abgeleiteten Objekttypen, in der Mengenlehre von Mengen und Teilmengen, in der Semantik von Begriffen und Unterbegriffen.

Wir möchten darauf hinweisen, dass ein Unterbegriff nie das reale Objekt sein kann. Genauso wenig ist in der OOP eine Instanz eines Objekttyps ein abgeleiteter Objekttyp.

Klassen kann man sich als Schubladen vorstellen, in welche man die zugehörigen Instanzen versorgen kann. Irgendwo muss es jetzt eine Vorschrift geben welche bestimmt, was alles in eine bestimmte Schublade gehört; diese Vorschrift entspricht der Definition der Klasse⁴.

Bleiben wir bei diesem Bild: Wir stellen uns einen Schubladenstock vor mit einer Schublade für Nägel und eine andere für Schrauben. Es könnte aber auch sein, dass es nur eine Schublade gibt für Kleinteile; dort würden dann sowohl Schrauben als auch Nägel versorgt.

Ähnlich verhält es sich mit den Klassen. Je nach Bedarf werden die Klassen enger oder weiter gefasst: Herr Muster möchte beispielsweise seinen Unfall in der Verwandtschaft nicht breit schlagen und spricht nur von einem „Herzproblem“. Anders ausgedrückt wählt er bewusst eine sehr breite Klasse und lässt damit vieles offen.

Bei der Beschreibung von Klassifikationen wird in Zusammenhang mit der Klassenbreite häufig von „Granularität“ gesprochen, wobei mit „feinerer Granularität“ weniger breite Klassen gemeint sind. Wie das obige Beispiel zeigt, bestimmt das Ziel der Anwendung wesentlich die benötigte Granularität.

zeichnungen verwendet werden.

⁴ Wenn die Klasse mit der Aufzählung ihrer Unterklassen oder Instanzen definiert wird, nennt man das „extensional“, wenn die Klasse mit einer Vorschrift (z.B. „alle Krankheiten, die das Herz betreffen“) definiert wird, nennt man das „intensional“.

4.3.3 Codes

Bleiben wir beim Beispiel der Schubladen: In einem Schubladenstock werden die einzelnen Schubladen zweckmässigerweise angeschrieben: Im Alltag wird man meist eine möglichst aussagekräftige Aufschrift wählen. Es ist aber auch denkbar, jede Schublade mit einer beliebigen Zeichenfolge zu versehen und beispielsweise ein Verzeichnis mit der Beschreibung jeden Schubladeninhaltes bereit zu stellen. Wichtig ist in diesem Fall, dass die gleiche Zeichenfolge nur auf genau einer Schublade vorkommt; einem sinnvollen Text braucht die Zeichenfolge nicht zu entsprechen.

Auch bei elektronischen Informationssystemen verwendet man Codes; diese können - müssen aber nicht - für den Menschen verständlich sein. Im ersten Fall spricht man von sprechenden Codes.

Codes werden in Datenbanken, Programmen etc. angewendet und in der Kommunikation zwischen Systemen; in Zusammenhang mit semantischen Standards interessiert vor allem letzteres. In guten elektronischen Informationssystemen kommt der Endbenutzer mit Codes nie in Berührung. In der Diskussion unter Fachleuten des Gesundheitswesens geht dies häufig vergessen und es werden hitzige Diskussionen über die Ausgestaltung sprechender Codes geführt. Diese Diskussionen sind unnötig: Gute Informationssysteme zeigen dem Benutzer anstelle des Codes z. B. einen Display Text⁵. Dies hat mehrere Vorteile, z. B. können Applikationen so leicht an verschiedene Sprachen angepasst werden.

4.3.4 Translations sind unvermeidbar

In den vorangehenden Abschnitten haben wir gesehen, dass die Terminologie (und damit auch Codes) und die verwendeten Informationsmodelle in der Realität nicht unabhängig sind. Applikationen im Gesundheitswesen (z. elektronische Patientendossiers) verwenden intern je nach Hersteller unterschiedliche Informationsmodelle (und damit unterschiedliche Datenstrukturen); intern verwenden sie ev. auch unterschiedliche Codes. Informationsmodelle und intern verwendete Codes können - müssen aber vom Hersteller nicht zwingend offen gelegt werden.

Die Kommunikation zwischen Systemen muss bestimmten Regeln folgen; sie hat also eine definierte Struktur (definierte Syntaktik, definiertes Informationsmodell und definierte Terminologie/Codierung).

Tim Benson benennt dies als Wire Format und stellt fest: „Any Interchange typically involves two translations: first from the native language of the sender to the wire format; second from the wire format to the native language of the recipient“⁶. Für Wire Format hat sich im deutschen Sprachraum „Austauschformat“ eingebürgert.

Auch wenn das Wire Format in Syntaktik und Semantik genau spezifiziert ist, bleiben die Translations bei Sender und Empfänger bestehen. Bei komplexen Sachverhalten sind diese Translations anfällig für Fehler; problematisch ist vor allem auch, dass – bei nicht dokumentierten internen Datenmodellen – ebenfalls die Translation eine Blackbox darstellen.

⁵ Denkbar sind auch andere Repräsentationen, z.B eine Grafik.

⁶ Benson T. Principles of Health Interoperability HL7 and SNOMED. 2nd ed. London: Springer; 2010

5 Semantische Grundlagen

5.1 Uneinheitliche Bezeichnungen

Die Diskussion der semantischen Interoperabilität leidet unter einer uneinheitlichen Verwendung der Fachausdrücke; dies ist ebenfalls in der einschlägigen Literatur der Fall. Eine Begriffsklärung ist deshalb unverzichtbar.

Je nach Kontext und Schule werden unterschiedliche Ausdrücke verwendet und es besteht die Gefahr, sich in akademischen Diskussionen um die richtigen Bezeichnungen zu verzetteln. Hinter den uneinheitlichen Bezeichnungen verstecken sich unterschiedliche Sichtweisen und deshalb stellt sich uns als erstes die Frage, welchen Gesichtspunkt wir für die Ordnung der Bezeichnungen als den wichtigsten erachten.

Wir haben uns entschlossen, die Bezeichnungen aus funktionaler Sicht zu ordnen. Dies erscheint uns deshalb sinnvoll, weil wir uns mit der semantischen Interoperabilität und mit Mapping beschäftigen. Es stellt sich insbesondere die Frage, wie sich verschiedene Ordnungssysteme funktional unterscheiden und wie gut sie ineinander überführbar sind.

Wir unterscheiden und definieren deshalb als erstes die Ordnungssysteme nach ihrer Funktion. Im Anhang findet sich eine Tabelle, in die hier verwendeten Ausdrücke mit alternativen, in der Literatur verwendeten Bezeichnungen verglichen werden.

5.2 Semiotisches Dreieck

Semantik beschäftigt sich mit der Bedeutung von Zeichen, Daten, Wörtern usw. „Bedeutung“ ist dabei in seiner einfachsten Form zu verstehen im Sinne von „Erkennen des Wortinhaltes“.

Eine gute Grundlage für die Diskussion und das Verständnis semantischer Zusammenhänge ist das Semiotische Dreieck⁷; in Abbildung 1 ist es grafisch dargestellt.

⁷ Das Semiotische Dreieck hat eine lange Geschichte: Bereits Platon hat ein Wort-Gegenstand-Modell zwischen Namen (Zeichen), Idee (Begriff) und Ding verwendet.

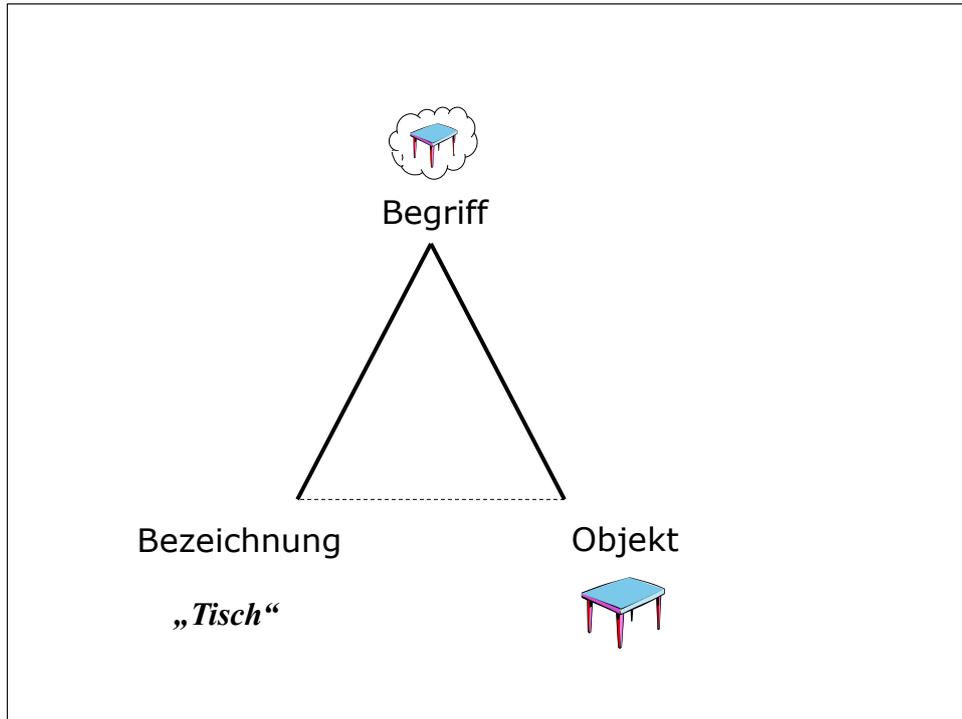


Abbildung 1: Semiotisches Dreieck

Im Semiotischen Dreieck wird streng zwischen Bezeichnung und ihrer Bedeutung unterschieden.

Von einem Objekt (im Dreieck rechts unten) machen wir uns eine bestimmte Vorstellung in unserem Kopf (im Dreieck die Spitze oben). Um diese Vorstellung anderen mitzuteilen, verwenden wir Bezeichnungen (im Dreieck links unten).

Der Begriff an der Spitze des semiotischen Dreiecks steht somit für unsere Vorstellungen von bestimmten Objekten und auch für die Bedeutung von Bezeichnungen oder Zeichen⁸. Der Begriff repräsentiert die Welt der Semantik.

Die drei Ecken des Dreiecks werden in der Literatur unterschiedlich benannt, z. B.:

- Bezeichnung: Ausdruck, Wort, Code, Bit, Term, Signifikant
- Begriff: Konzept, Concept, Signifikat
- Objekt: Denotat, Object

Wichtiger als die Bezeichnungen der drei Ecken ist das Verständnis ihrer Unterschiede:

Die drei Ecken des semiotischen Dreiecks repräsentieren drei unterschiedliche Welten: Rechts befind-

⁸ Der Ausdruck „Zeichen“ kann weiter differenziert werden und natürliche und künstliche, d.h. von Menschen geschaffene, umfassen. Letztere sind diejenigen, die in unserem Zusammenhang interessieren. Wörter sind eine Untermenge dieser kulturell erschaffenen Zeichen. Es ist Definitionssache, ob wir unter Wörtern nur Buchstabenstrings verstehen oder auch komplexere Zeichen, die auch Zahlen und mehr enthalten können. Wörter können auch als Terme bezeichnet werden. Sicher gehören Codes auch zu diesen Zeichen. In unserem Zusammenhang möchte ich alle diese Art Symbole oder Zeichen der einen Ecke (links unten) des Dreiecks zuordnen, insofern es sich bei Wörtern, Codes und Termen immer um äussere Ausdrücke handelt, die wir verwenden um zu kommunizieren

det sich die Realwelt, in der die Gesetze der Realwelt (Physik, Chemie, Medizin, etc.) gelten⁹. Links ist die die Welt der Sprache, in der die Gesetze der jeweiligen Sprache (Syntaktik usw.) gelten. Oben befindet sich die Welt der Begriffe, also der Vorstellungen, die wir von bestimmten Objekten haben. Genauso wie die Gesetze der Sprache nicht die Gesetze der Physik sind, sind auch die Gesetze der Semantik nicht diejenigen der Physik, noch diejenigen der Sprache, sondern eigene. Die Elemente der drei Welten (Bezeichnung, Begriff, Objekt) können entsprechend auch nicht einander gleichgesetzt werden. Auch wenn sie sich in einem konkreten Fall aufeinander beziehen, herrschen doch zwischen ihnen keine 1:1 Beziehungen.

a) **Bezeichnung ≠ Begriff:** Der Begriff *Tisch* kann mit dem deutschen Wort "Tisch", aber auch mit dem französischen Wort „table“ bezeichnet werden. Das englische Wort „table“ steht nicht nur für den Begriff *Tisch*, sondern auch für den Begriff, der mit dem deutschen Wort "Tabelle" bezeichnet wird.

a) **Begriff ≠ Objekt:** Ein Begriff *Tisch* gilt für viele Einzelobjekte. Ein Einzelobjekt kann unter unterschiedlichen Begriffen verstanden werden, ein Objekt „Tisch“ kann zum Beispiel unter dem Begriff *Möbel, Gegenstand, Heizmaterial, Unterlage* usw. verstanden werden.

c) **Objekt ≠ Bezeichnung:** Zwischen dem Objekt und der Bezeichnung gibt es **keine direkte** Relation, sondern die Beziehung läuft immer **über das Verständnis im Kopf** von Sprecher und Zuhörer, also über die Begriffe in ihren Köpfen. Die Basislinie im Dreieck wird aus diesem Grund gestrichelt gezeichnet. Wenn ich keine Vorstellung von einer Bezeichnung habe, kann ich diese auch nicht sinnvoll verwenden und mit einem Objekt verbinden.

Die oben dargestellten Unterscheidungen sind keinesfalls bloss theoretische, sondern haben praktische Konsequenzen: Für eine Semantische Interoperabilität darf sich die Kommunikation nicht nur auf Bezeichnungen abstützen; es muss sichergestellt sein, dass unter einer bestimmten Bezeichnung auch der gleiche Begriff verstanden wird: Wenn sich die Kommunikation ausschliesslich auf Bezeichnungen stützt, ist sie anfällig für Mehrdeutigkeiten (Homonymie: der selbe sprachliche Ausdruck steht für verschiedene Begriffe).

5.2.1 Eigenschaften der Begriffe, die bei Wörtern nicht vorkommen

Für Begriffe gelten **andere Regeln als für Bezeichnungen**. Letztere folgen der Grammatik. Bei Begriffen hingegen sind nicht die Regeln der Grammatik, sondern die Phänomene der Implikation und der Kontextabhängigkeit zu beachten:

Implikation:

- a) Ein Begriff impliziert seinen Oberbegriff (Radius -> Knochen).
- b) Ein Begriff impliziert einen assoziierten Begriff (Radius -> Unterarm)

Kontextabhängigkeit:

- a) Bruch, Humerus -> Fraktur
- b) Bruch, Leiste links -> Hernie

⁹ Als Objektwelt kommt allerdings auch eine rein imaginierte Welt in Frage. Auch in diesem Fall repräsentiert die Spitze links unten im Dreieck die Welt auf die sich die Gedanken (Spitze oben), bzw. die Bezeichnungen (Spitze links) beziehen. Die Objektwelt bleibt in diesem Sinn die Objektwelt, ob sie real ist oder nicht.

Implikation und Kontextabhängigkeit zeigen eine weitere wichtige Eigenschaft der Begriffe, nämlich der Tatsache, dass Begriffe einander durchdringen. Dies hat damit zu tun, dass Begriffe gedankliche Felder sind und nicht eindeutig lokalisierte Punkte. Das Feld mit der Bedeutung *Ulna* und das Feld mit der Bedeutung *Knochen* durchdringen einander. Die zugehörigen Wörter tun dies nicht¹⁰. Die Phänomene der Implikation, der Kontextabhängigkeit und der Durchdringung spielen in der Semantik eine grosse Rolle.

Insbesondere spielt es eine grosse Rolle, auf welche Weise Begriffe einander durchdringen. Die Art der Durchdringung hat viel mit der gegenseitigen Anordnung der Begriffe zu tun. Hier unterscheiden wir verschiedene Arten der Begriffsanordnung oder der Begriffsarchitektur (siehe unten). Ordnungssysteme wie ICD-10, CHOP, SNOMED basieren auf jeweils einem bestimmten Typ von Begriffsarchitektur, der letztendlich entscheidend für ihre Ausdrucksmöglichkeiten ist. Die vorhandenen (oder nicht vorhandenen) Ausdrucksmöglichkeiten spielen beim Mapping und bei der Interoperabilität die entscheidende Rolle.

5.2.2 Kommunikation

Ein Datenaustausch findet statt, wenn die Kommunikationsteilnehmer die gleichen Bezeichnungen verwenden, ein Informationsaustausch jedoch erst dann, wenn sie diese auch gleich verstehen, d.h. unter der verwendeten Bezeichnung auch den gleichen Begriff verstehen.

Für einfache Botschaften mögen die Daten als identisch mit der Information erscheinen: Je einfacher und je konkreter die Botschaft ist, umso selbstverständlicher ist auch die Interpretation. Doch je schwieriger, komplexer und abstrakter der Inhalt ist, umso schwieriger wird ein gemeinsames Verständnis.

5.3 Ordnungssysteme, Klassifikationen, Codesysteme, Terminologien usw.

5.3.1 Vorbemerkung

Die Begriffe eines Fachgebietes können unterschiedlich angeordnet werden und die Art der Anordnung (Semantische Architektur) spielt eine wesentliche Rolle für das Verständnis und die Arbeit im Fachgebiet. Die Art der Anordnung ist aber auch entscheidend für die semantische Interoperabilität, fürs Mapping usw. Die Ordnungssysteme sollen deshalb an dieser Stelle nach der Art ihrer inneren Struktur differenziert werden.

5.3.2 Ordnungssysteme ohne explizite innere Struktur

Bei solchen Systemen wird kein Wert auf die Beziehungen (Relationen) zwischen den einzelnen Begriffen gelegt. Das Spektrum reicht von der einfachen Aufzählung der verwendbaren Ausdrücke bis zu Systemen, in denen diese genau definiert sind. In diesem Dokument verwenden wir für ein solches Benennungssystem ohne explizite innere Struktur auch die Bezeichnung „Terminologie“¹¹.

¹⁰ Die Bezeichnungen „Ulna“ und „Knochen“ durchdringen einander nicht, sie haben allenfalls einen gemeinsamen Buchstaben „n“. Allerdings können auch Bezeichnungen einander durchdringen. So steckt in der Bezeichnung „Knochen“ das Wort „noch“, im Wort „Wort“ das Wort „Wo“ und das Wort „Ort“. Diese Art Durchdringung ist allerdings rein zufällig und ohne Bedeutung.

¹¹ Die Bezeichnung „Terminologie“ wird von anderen Autoren auch für Ordnungssysteme im Allgemeinen verwendet.

Ein Benennungssystem ohne spezielle innere Struktur kann bereits Klassen enthalten. Auch wenn die Klassen noch nicht angeordnet sind, z.B. in Form eines Baumes, sondern einfach nebeneinander existieren, müssen sie bereits eine wichtige Forderung erfüllen: sie müssen disjunkt sein, d.h. sie dürfen sich nicht überschneiden. Sie können zwar eine sichtbare äussere Anordnung aufweisen, wie z.B. in eine Aufzählung in einem alphabetischen Verzeichnis, dabei handelt es sich aber um eine Äusserlichkeit. Die Semantik wird dabei nicht abgebildet.

Ein Beispiel für eine solche im Prinzip simple Aufzählung ist das alphabetische Verzeichnis der ICD-10. Das Verzeichnis ist rein alphabetisch aufgebaut, hat also keine semantische Struktur. Allerdings verweist jeder Eintrag im Hintergrund auf Stellen im hierarchischen, d.h. semantisch sehr wohl strukturierten systematischen ICD-10 Verzeichnis.

Noch deutlicher verhält es sich bei SNOMED CT. Auch hier existiert eine Aufzählung, nämlich die Aufzählung der klinischen Terme (CT=clinical terms). Diese Aufzählung ist zwar semantisch unstrukturiert, weist aber auf den hoch strukturierten inneren Kern von SNOMED hin, (Dieser ist höher strukturiert als die Kernstruktur der ICD-10). SNOMED nennt die aufzählende Oberfläche präkoordiniert und den Kern postkoordiniert¹².

Die Verbindung von Aufzählungsoberfläche mit einem Ontologiekern wie bei SNOMED bietet unbestreitbare Vorteile, indem die Vorteile beider Ordnungssysteme kombiniert werden. Was es braucht, damit diese Vorteile auch in der Praxis nutzbar sind, wird in Kapitel 6 erläutert.

5.3.3 Hierarchische Ordnungssysteme

Hierarchische Ordnungssysteme ordnen die Begriffe in einem Hierarchiebaum. Die differenziertesten Begriffe bilden die Blätter des Baumes. Mehrere Blätter sind als Unterbegriffe einem Oberbegriff untergeordnet. Dieser kann einem weiteren Oberbegriff untergeordnet werden, bis über mehrere Stufen die Wurzel (der allgemeinste Begriff des Baumes) erreicht ist. Die hierarchische Struktur ist einfach, klar und nützlich (solange sie mit der Realität übereinstimmt).

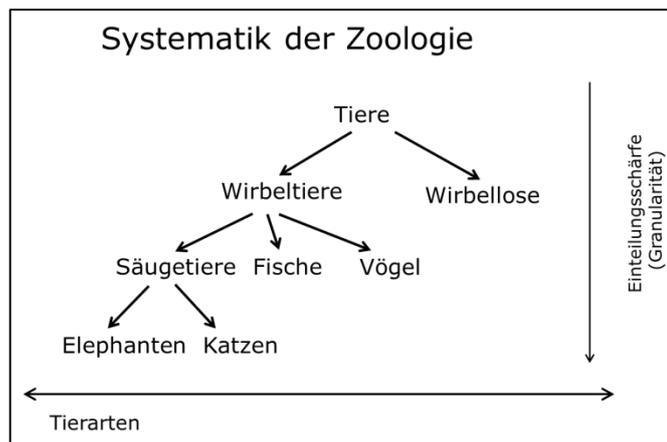


Abbildung 2: Beispiel eines hierarchischen Baumes¹³

5.3.4 Semantische Achse oder Dimension, semantischer Freiheitsgrad

Der Begriff der semantischen Achse oder Dimension, bzw. des semantischen Freiheitsgrades ist we-

¹² Es handelt sich um eine mehrdimensionale Struktur (siehe Kapitel 5.3.6)

¹³ Es handelt sich um eine Skizze. Natürlich enthält der Baum der Zoologie viel mehr Tierarten und Klassen.

sentlich für das Verständnis der unterschiedlichen Ordnungssysteme und für das Verständnis von semantischen Repräsentationen und Zuordnungen, von Aussagen, Mappings etc. Der Begriff kann anhand des hierarchischen Systems einfach erklärt werden.

Wenn man ein hierarchisches System, d.h. einen Baum, aufzeichnet, erstreckt sich die Darstellung über zwei Dimensionen (siehe Abbildung 2). Zuunterst findet man die "Blätter" des Baumes¹⁴. Wenn man diese Blätter auf einer Geraden anordnet, erhält man in der Horizontalen die semantische Grundachse dieser Struktur. Die zweite, vertikale Achse erstreckt sich von der (oben gezeichneten) Wurzel des Baumes nach unten zu den Blättern. Während sich auf der Grundachse von Blatt zu Blatt die Bedeutung (Semantik) ändert, wächst auf der zweiten Achse nur die Granularität der Begriffe. Mit anderen Worten: Die Blätter enthalten feingranulär die Details des Systems, während die Begriffe darüber immer gröber (grobgranulärer) werden, je näher sie zur Wurzel kommen. So ist in Abbildung 2 eine Katze eine präzisere, feingranulärere Bezeichnung als Säugetier, informatisch also das feinere Korn. Semantisch kommt beim Übergang von Katze zu Säugetier nichts dazu, was nicht schon in der Katze enthalten war, beim umgekehrten Übergang vom Säugetier zur Katze kommen zwar zusätzliche Informationen hinzu, allerdings nur im Sinn der zusätzlichen Spezifikation und nicht einer Änderung – wie sie zum Beispiel beim Übergang von der Katze zum Elefanten stattfinden würde. Deshalb sehen wir die vertikale Achse einer Hierarchie als Granularitätsachse, auf der die Semantik nicht ändert, nur präziser oder gröber wird, die horizontale Achse hingegen, auf der die Information auch ihre Qualität ändert, als die eigentliche semantische Achse. Der typische Klassifikationsbaum wie in Abbildung 2 entspricht somit semantisch einer Dimension oder Achse. Weil nur einmal gewählt werden kann (nämlich beim Blatt), sprechen wir von einem semantischen Freiheitsgrad.

5.3.5 Unidirektionalität, Multihierarchie

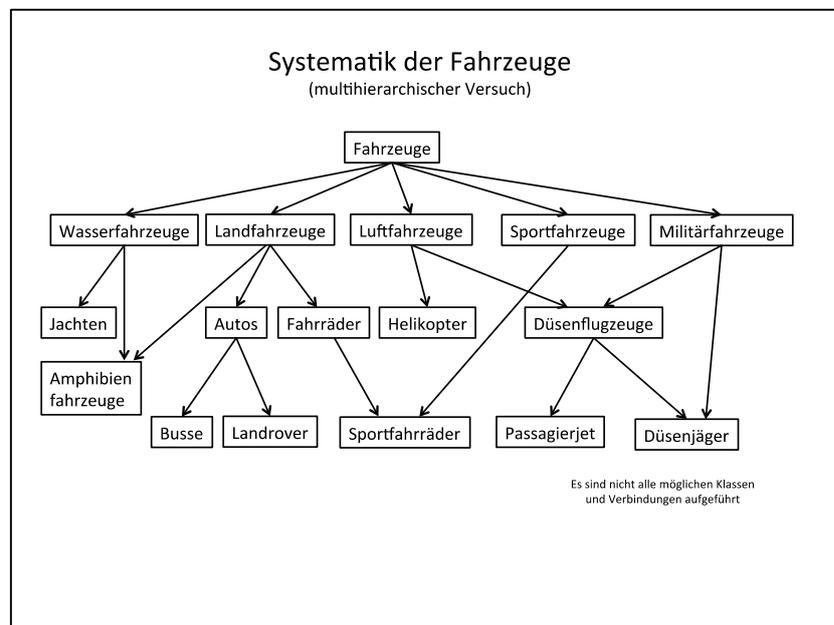


Abbildung 3: Beispiel eines multihierarchischen Baumes¹⁵

¹⁴ Dies erscheint auf den ersten Blick paradox, da sich bei einem Baum die Blätter natürlich oben und die Wurzel unten befinden. Die quasi inverse Darstellung wie in Abbildung 2 hat sich allerdings eingebürgert. Sie entspricht der Pyramide, wie wir sie z.B. in der Befehlsstruktur einer Armee finden. Diese menschengemachte Hierarchie hat auch eine Baumstruktur, und in ihr finden sich der General oben und die Soldaten unten.

¹⁵ Die Klassen Luftfahrzeuge und Militärfahrzeuge überschneiden sich z.B. bei den Düsenjägern. Die Klassen Luftfahrzeuge und Militärfahrzeuge sind somit „nicht disjunkt“.

Multihierarchien sind Hierarchien mit Überschneidungen, d.h. Klassen auf der gleichen Hierarchiestufe überschneiden sich und sind somit nicht mehr „disjunkt“.

Neben der Disjunktheit (zwischen Klassen auf der gleichen Ebene) ist bei hierarchischen Ordnungssystemen eine weitere Eigenschaft wichtig: die Unidirektionalität (Gerichtetheit). Bei einem Baum im strengen Sinn darf ein Unterbegriff nur einen Oberbegriff haben. Das heisst, dass alle Verzweigungen in einem Baum nur in eine Richtung gehen dürfen. Wenn diese Bedingung *nicht* erfüllt ist, d.h. wenn ein Begriff mehrere Oberbegriffe haben darf, spricht man von einer Polyhierarchie oder Multihierarchie. Eine Multihierarchie ist kein Baum mehr, sondern enthält auch netzartige Verzweigungen. Die Blätter bilden zwar wieder eine Achse wie bei Bäumen, doch zwischendurch ist die Disjunktheit gebrochen, d.h. auf dem Weg vom Blatt zur Wurzel sind alternative Wege möglich: eine Hepatitis A ist sowohl eine Viruskrankheit als auch eine Leberkrankheit.

Beim Versuch, Krankheiten in einem Baum anzuordnen, gerät man schnell in Versuchung Multihierarchien zu bauen – ganz im Gegensatz zu den Verhältnissen beim Versuch, Tierarten zu ordnen. Das ist kein Zufall, sondern hat mit der engen Verwandtschaft aller Tiere zu tun, die sich während der Evolution aus der gleichen Wurzel heraus differenziert haben. Deshalb ordnen sich die Tierarten natürlicherweise unidirektional (Die Differenzierung der Arten erfolgt parallel zum unidirektionalen Zeitstrahl).

Die Disjunktheit der Tierarten wiederum erklärt sich durch die Artengrenzen. Dort wo sich Tierarten auch mischen können (Maulesel) ist die Disjunktheit gebrochen und die Baumhierarchie wird zur Multihierarchie. Das gleiche stellen wir innerhalb der Arten bei den Rassen fest. Bei den medizinischen Diagnosen sind die Verhältnisse schon von Beginn an viel komplexer, sodass eine klassische Baumanordnung von allen Diagnosen immer ein künstliches Konstrukt ist (siehe ICD, CHOP usw.).

Multihierarchien sind somit keine Bäume, denn ihre Klassen sind nicht disjunkt und ihre Verzweigungen sind nicht unidirektional, sondern bilden Netzstrukturen. Multihierarchien sind jedoch weiterhin Systeme mit nur einer semantischen Dimension, wie es auch die klassischen Hierarchieebenen sind. In dieser Hinsicht unterscheiden sie sich klar von den mehrdimensionalen Ordnungssystemen (vgl. 5.3.6).

5.3.6 Mehrdimensionale Ordnungssysteme

Eine mehrdimensionalen Ordnungssystem kombiniert mehrere Bäume: Jeder Baum entspricht einem Freiheitsgrad, bzw. einer Dimension oder Achse (siehe oben). Die Achsen sind hier die Grundachsen der jeweiligen Hierarchieebenen. Sie können als Koordinatenachsen eines Koordinatensystems angesehen werden.

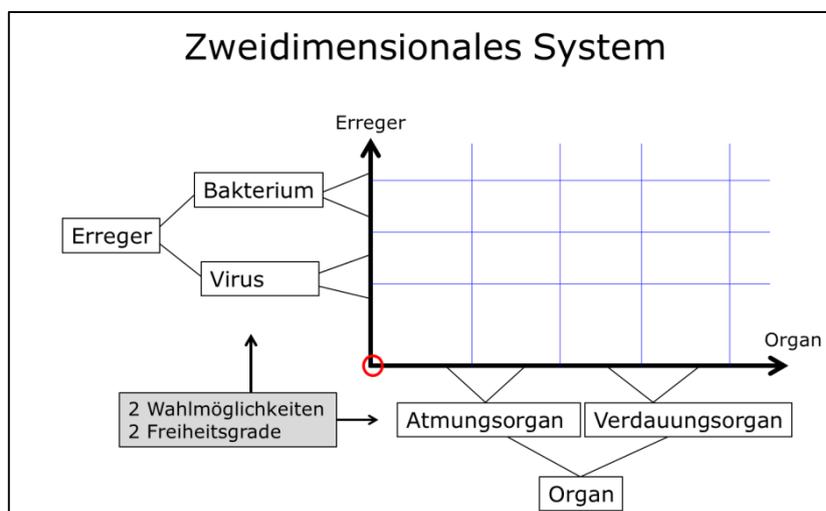


Abbildung 4: Mehrdimensionales System

Mehrdimensionale Ordnungssysteme bieten den Vorteil der Kombinatorik: Während sich die Zahl der Begriffe (Werte) auf den einzelnen Achsen zur Gesamtzahl der Begriffe addieren, multipliziert sich die Zahl der mit den Werten auf mehreren Achsen erreichbaren kombinierten Begriffe. So können mit $a+b+c$ einfachen Begriffen $a*b*c$ kombinierte Begriffe definiert werden. $a*b*c$ ist dabei immer grösser als $a+b+c$, was die Potenz der mehrdimensionalen Systeme ausmacht. So können bei den Verhältnissen von Abbildung 4 mit 1000 Erregern und 20 Organen, d.h. mit 1020 unterschiedlichen Begriffen 20'000 Felder, d.h. unterschiedliche Kombinationen definiert werden.

Ein Beispiel für ein mehrdimensionales Ordnungssystem im pflegerischen Bereich ist die ICNP¹⁶.

Fazit: Mehrdimensionale Ordnungssysteme sind eindimensionalen (Bäumen) bezüglich Ausdrucksstärke überlegen. Sie sind effizienter, d.h. mit weniger Klassen (Codes) sind mehr Differenzierungen (Schubladen) möglich. Dies wird dadurch erkauft, dass pro Einzelfall mehrere Codes angegeben werden müssen.

5.3.7 Multifokale Ordnungssysteme

Multidimensionale Ordnungssysteme kennen zwei systematische Probleme: Ihre Effizienz nimmt zu (siehe oben), je mehr Dimensionen (Freiheitsgrade) sie aufweisen. Mit der Zunahme der Dimensionen werden sie jedoch immer unübersichtlicher – wer kann sich schon sieben Dimensionen räumlich vorstellen? Gleichzeitig nehmen im Koordinatensystem, das durch die Achsen aufgefaltet wird, die Punkte zu, welche keiner sinnvollen Kombination mehr entsprechen. Durch Kombination der Werte auf den einzelnen Achsen könnte z.B. ein "posttraumatischer offener Diabetes des Muskelgewebes" definiert werden; etwas, das in der Realität nicht existiert. Der Overkill an Kombinationen rührt daher, dass sich bei klassischen Koordinatensystemen alle Achsen in einem Nullpunkt schneiden und deshalb immer alle sichtbar sind. Bei multifokalen Ordnungssystemen ist das nicht mehr der Fall, ihre Überlegenheit entsteht dadurch, dass die Achsen (Freiheitsgrade) gekapselt werden. Dadurch sind nur noch diejenigen Achsen sichtbar, welche durch Werte auf übergeordneten Freiheitsgraden geöffnet werden. Auf diese Weise reduziert sich für einen aktuell vorliegenden Fall die Zahl der Freiheitsgrade und auch ein System mit mehreren tausend Freiheitsgraden bleibt völlig übersichtlich.

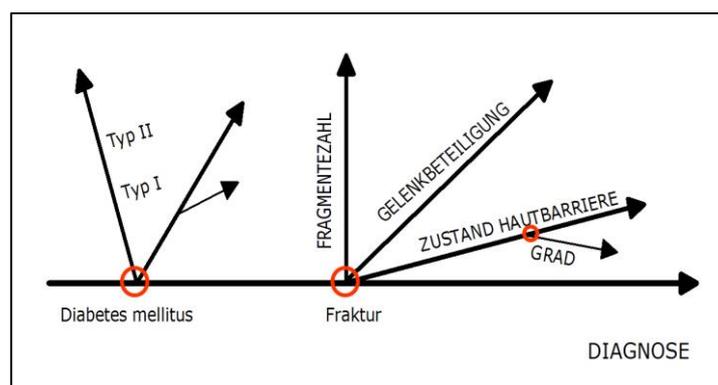


Abbildung 5: Multifokales System

In Abbildung 5 sind die Kapselungen (Fokusse) mit roten Kreisen gekennzeichnet. So ist der Freiheitsgrad der Fragmentezahl dem Freiheitsgrad Diagnose untergeordnet und mit weiteren Freiheitsgraden nur sichtbar, wenn auf der Achse Diagnose der Wert Fraktur aktiviert ist.

¹⁶ ICNP = International Classification of Nursing Practice.

5.3.8 Ontologien

„Ontologie“ wird in der Literatur sehr unterschiedlich definiert¹⁷. Im Bereich der Computerlinguistik wird gerne von „formalen Ontologien“ gesprochen, wobei „formal“ ausdrückt, dass nach Regeln der mathematischen Logik vorgegangen wird. Generell wird angenommen, dass Ontologien Begriffe eindeutig anordnen und dabei unterschiedliche, aber wohldefinierte Relationen zwischen den Begriffen verwenden (relationships oder link types).

Das heisst, dass neben den beiden hierarchischen Relationen, wie sie in hierarchischen Bäumen zwischen den Klassen vorkommen (is-a und is-part-of), Ontologien weitere und unterschiedliche Qualitäten von Relationen enthalten, welche die Beziehung zwischen den Begriffen abbilden. Diese Relationen sind u.E. notwendige Elemente, jedoch keine Besonderheit der Ontologien, sondern können bereits mit multifokalen Ordnungssystemen ausgedrückt werden. Ontologien ermöglichen aber darüber hinaus das Ziehen von Schlüssen mit Hilfe diese Relationen. Was Ontologien somit kennzeichnet, ist ihr Fähigkeit zwischen verschiedenen Freiheitsgraden Schlüsse zu ziehen. In diesem Sinn verwenden wir im vorliegenden Dokument die Bezeichnung „Ontologie“ und grenzen ihn von anderen multidimensionalen oder multifokalen Ordnungssystemen ab.

SNOMED CT und die Ontologien des Semantic Web sind Ontologien im oben beschriebenen Sinn.

Alle Ontologien sind Beschreibungen eines Ausschnittes der Welt. Bei der Beschreibung wird oft grosser Wert darauf gelegt, dass sie nicht nur präzise, sondern auch korrekt und widerspruchsfrei ist, d.h. eine endgültige und widerspruchsfreie Repräsentation der Realität ermöglicht. Das spiegelt sich nicht zuletzt darin wieder, dass für das Aussagesystem einer solchen Ontologie eine strenge mathematische Logik (FOL¹⁸ oder DL¹⁹) anwendbar sein muss, welche das Postulat der Monotonie erfüllt. Das Postulat der Monotonie ist jedoch u.E. keine zwingende Bedingung für Ontologien.

¹⁷ Ursprünglich stammt der Ausdruck aus der Philosophie, wo er die Lehre des Seienden bezeichnet. Gegen Ende des 20. Jahrhunderts kam er in der Informatik auf, um auszudrücken, was mit Datenbankinhalten gemeint war. Philosophen und Informatiker verwenden den Ausdruck deshalb unterschiedlich. Wir nehmen in unserem Text aber nur Bezug auf die Verwendung in der Informatik.

¹⁸ FOL = First Order Logic = mathematische Logik ersten Grades: Klassische Form der mathematischen Logik, die streng **monoton** ist, was bedeutet, dass der Wahrheitswert einer Aussage nicht ändern darf, wenn neue Sätze zum System dazukommen. Bei nichtmonotoner Logik ist eine Änderung des Wahrheitsgehaltes erlaubt.

¹⁹ DL = Description Logic = Beschreibungslogik: Die DL ist eine Fragment aus der FOL, welches die bei der FOL heikle Entscheidbarkeit verbessert, jedoch die Expressivität dabei einschränkt. DL wird oft bei Ontologien und im Semantic Net eingesetzt. Sie ist wie FOL monoton. Beide Logiken sind deshalb für die praktische Aufgabe des NLP (Natural Language Processings) schlechter geeignet als nichtmonotone Logiken.

5.4 Semantische Analyse von Diagnoseinformationen

5.4.1 Beispiel Abdominalhernien ICD-10 WHO

Nachfolgend stellen wir dar, wie eine semantische Analyse eines Fachbereichs mit Hilfe der semantischen Achsen (semantischen Freiheitsgrade) durchgeführt werden kann.

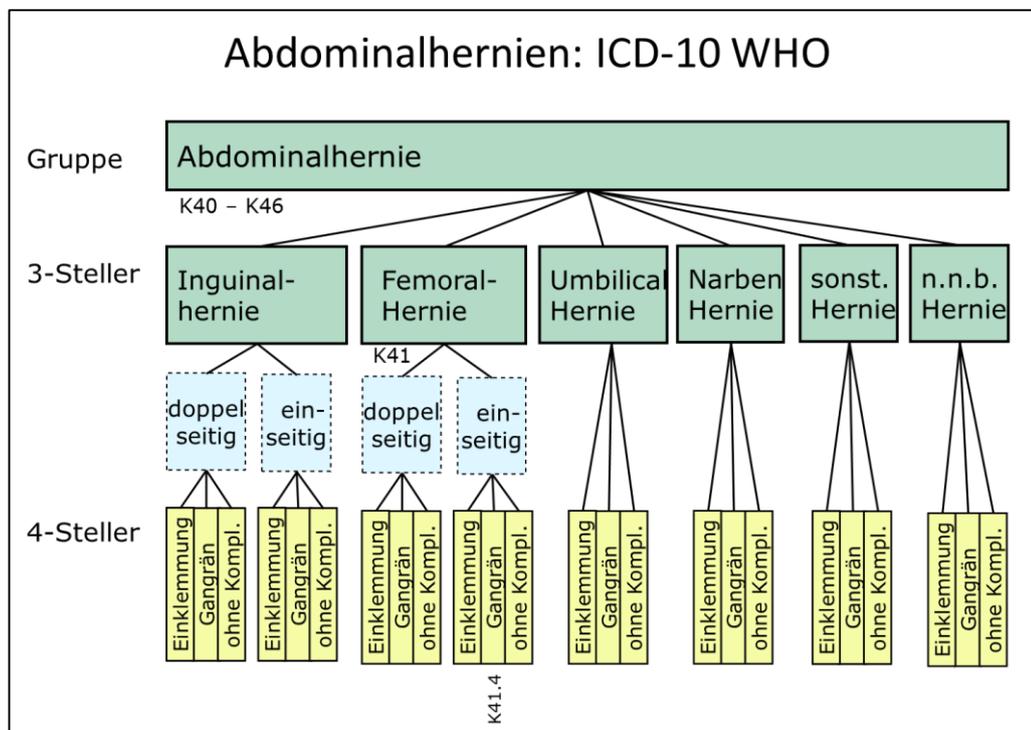


Abbildung 6: Semantische Analyse eines Abschnittes der ICD-10-WHO

In Abbildung 6 ist die Gruppe der Abdominalhernien mit ihren Ästen und Blättern dargestellt. Die ICD-10 ist in Form eines hierarchischen Baums aufgebaut²⁰. Von oben nach unten sind dabei die hierarchischen Stufen des Baums: Kapitel – (Unterkapitel) – Gruppe – 3-Steller – 4-Steller – 5-Steller.

Hinter der formalen Oberfläche, d.h. hinter der Baumstruktur versteckt sich die tatsächliche semantische Gliederung, welche die semantische Information repräsentiert, die mit Hilfe des ICD-10-Codes ausgedrückt wird. In der Gruppe der Abdominalhernien können wir drei unterschiedliche Freiheitsgrade erkennen; in Abbildung 6 sind sie mit den Farben grün, blau und gelb markiert. Jeder Freiheitsgrad bezeichnet eine Qualität²¹, die verschiedene Ausprägungen haben kann. Im vorliegenden Beispiel sind dies:

²⁰ Mit der Kreuz-Stern Kodierung (siehe ICD-Manual) führt die ICD-10 zwar zum primären Baum einen zweiten Freiheitsgrad ein. Diese Erweiterung ist allerdings nur für die wenige Teilbereiche des Diagnosebaums durchgeführt und zudem inkonsequent. So wird z.B. für die Masernpneumonie der zweite Freiheitsgrad verwendet, um Manifestation (Pneumonie) und Erreger (Masernvirus) getrennt zu kodieren (Kodes = B05.2+ plus J17.1*) – für andere Pneumonien hingegen verzichtet die ICD-10 auf den zweiten Freiheitsgrad, z.B. für die Pneumokokkenpneumonie (Kode = J13) und die meisten anderen Pneumonien. Allerdings wird schnell klar, dass zwei Freiheitsgrade nicht ausreichen und auch nicht konsequent durchgeführt werden können. Die US-amerikanische Überarbeitung der ICD-10, die ICD-10-CM, verzichtet gegenüber der ICD-10-WHO und der ICD-10-GM deshalb bereits konsequent auf die Kreuz-Stern-Kodierung.

²¹ Attribut

Tabelle 1: In den Abdominalhernien der ICD-10-WHO (siehe Abbildung 6) gefundene Freiheitsgrade

Attribut	Ausprägungen	Farbe
Lokalisation	Inguinal, femoral, umbilical, Narbe, sonstige, nicht näher bezeichnete	Grün
Seitigkeit	Doppelseitig, einseitig	Blau
Komplikationen	Einklemmung, Gangrän, keine	Gelb

Je nach Hierarchielevel, können verschiedene Informationen ausgedrückt werden: Auf dem Level der Gruppe (K40-K46) ist bekannt, dass es sich um eine Abdominalhernie handelt. Die genauere Information über die Lokalisation wird auf der Stufe 3-Steller gegeben, z.B. K41 = Femoralhernie. Auf der Stufe des 4-Stellers wird zusätzlich die Seitigkeit und die Komplikation spezifiziert, z.B. K41.4 = einseitig und ohne Komplikationen. Die Information über die Seitigkeit erfolgt dabei implizit, aber eindeutig: Wäre die Hernie nämlich doppelseitig, müsste statt K41.4 der Code K41.1 vergeben werden. Der 4-Steller trägt in diesem Fall neben der Information über die Komplikationen die Information über die Seitigkeit implizit und eindeutig mit. Dieses Verhalten führt dazu, dass eine Femoralhernie, von der man nichts über die Komplikationen weiss, von der jedoch bekannt ist, dass sie einseitig ist, nicht eindeutig codiert werden kann²².

Ein weiteres Problem ist, dass gewisse Informationen in der ICD-10 WHO überhaupt nicht codiert werden können. So kann z.B. nicht gesagt werden, ob es sich um eine Rezidivhernie handelt oder nicht. Ebenso wenig kann ausgedrückt werden, ob eine Inguinalhernie direkt oder indirekt ist. Den ersten Mangel korrigiert die deutsche Erweiterung der ICD-10 WHO, die ICD-10-GM, die mehr oder weniger unverändert in die Schweiz übernommen worden ist (siehe Abbildung 7).

5.4.2 Beispiel Abdominalhernien ICD-10 GM

Internationale Codes werden häufig lokal angepasst. So wurde ICD-10 WHO von DIMDI²³ für Deutschland angepasst und als ICD-10-GM²⁴ publiziert. Abbildung 7 zeigt die Erweiterungen, die das DIMDI für die Gruppe der Abdominalhernien durchgeführt hat.

²² Man muss in diesem Fall entweder den 3-Steller ausgeben, was aber in gewissen Situationen verboten ist, da der 3-Steller hier kein Endcode ist, oder den 4-Steller K41.9 angeben. Letzteres ist in der Tat der korrekte Code. Dieser Code wird aber auch vergeben um auszudrücken, dass die Femoralhernie KEINE Komplikationen hat. Mit anderen Worten: Bei K41.9 geht die Information über die Komplikation verloren, und man weiss nur sicher dass die Hernie einseitig ist.

²³ Deutsches Institut für Dokumentation und Information.

²⁴ GM = German Modification

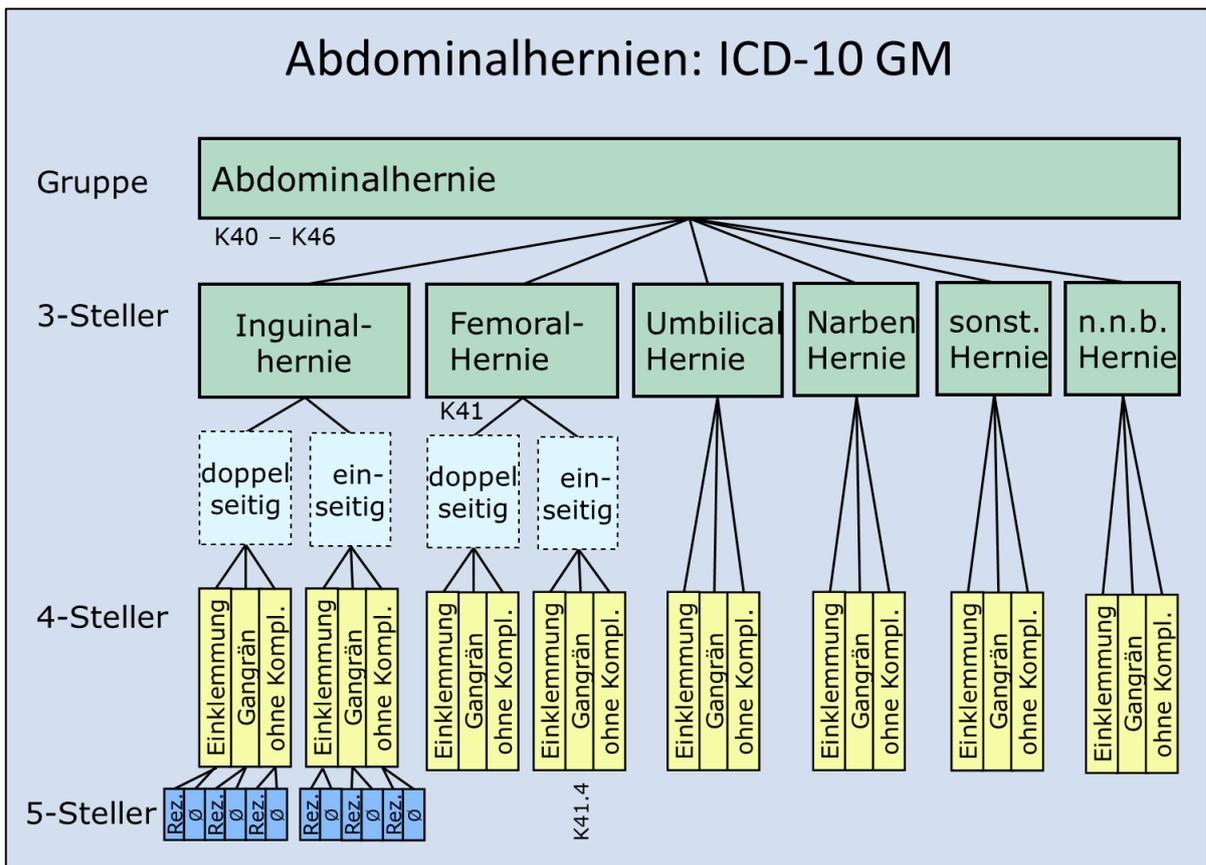


Abbildung 7: Gleicher Ausschnitt wie in Abbildung 6, jetzt aber in der Version ICD-10-GM

In Abbildung 7 repräsentieren die blauen Kästchen einen zusätzlichen Freiheitsgrad, der die Information darüber enthält, ob es sich um ein Rezidiv einer Hernie handelt oder nicht. Diese Information kann also in der ICD-10 GM im Vergleich zur ICD-10 WHO neu repräsentiert werden. Dies ist aber nur für die Inguinalhernien möglich. Für die übrigen Hernien kann diese Information nicht ausgedrückt werden. Die Information, ob eine Inguinalhernie direkt oder indirekt ist, kann wie in der ICD-10 WHO auch in der ICD-10 GM nicht ausgedrückt werden.

5.4.3 Multifokale Darstellung der semantischen Achsen

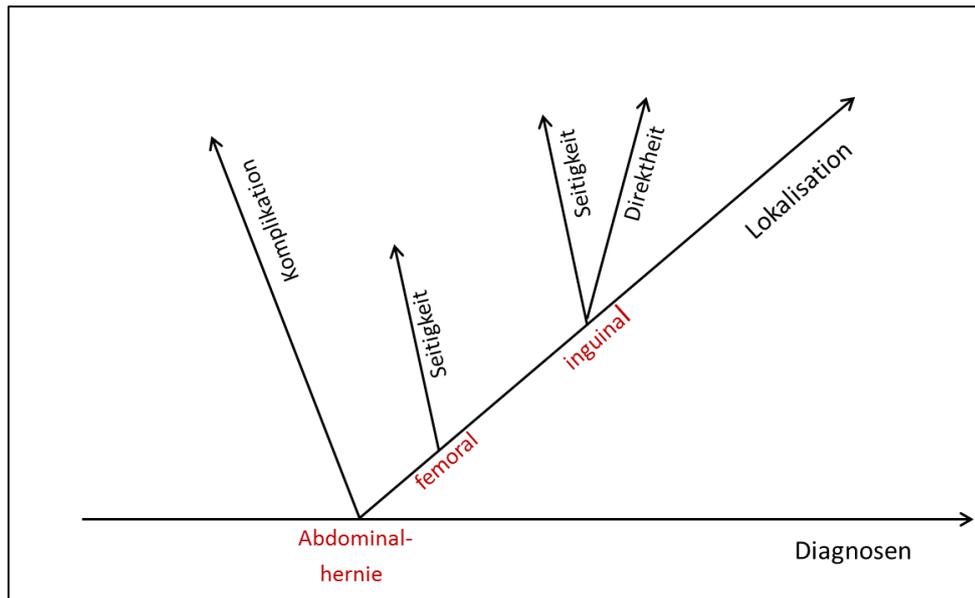


Abbildung 8: Darstellung des multifokalen Achsengerüsts des Ausschnitts von Abbildung 7 (Die Namen der Achsen sind schwarz, die Namen der Fokus-Knoten rot)

Abbildung 8 zeigt das Achsengerüst, welches aus der semantischen Analyse des Baums der ICD-10-GM folgt. Die Freiheitsgrade (Achsen) sind schwarz bezeichnet, während die Werte, von denen die untergeordneten Freiheitsgrade abgehen, rot bezeichnet sind. Bei einer Hierarchie bilden die Werte einen Baum, während bei der multifokalen Architektur die Freiheitsgrade einen Baum bilden. Das so für die Abdominalhernien entstehende multifokale Achsengerüst enthält 5 Freiheitsgrade, während der Baum der ICD-10 nur einen Freiheitsgrad erlaubt. Die Probleme, die sich aus dieser Vereinfachung ergeben, sind in Kapitel 5.5 näher erklärt.

5.4.4 Abbildung von Inhalten auf die verschiedenen Typen von Ordnungssystemen

In Kapitel 5.3. zeigten wir, wie weit unterschiedliche Typen von Ordnungssystemen in der Lage sind komplexe Semantiken präzise abzubilden. Sie unterscheiden sich vor allem in der Anzahl Dimensionen/Freiheitsgrade, die sie getrennt voneinander repräsentiert und behandeln können:

- a) Hierarchiebaum 1
- b) Mehrdimensionales System 2 – ca. 20
- c) Multifokales System 10 – 1000 und mehr

Wenn wir Freiheitsgrade als Achsen eines geometrischen Systems verstehen, spannen diese einen „**semantischen Raum**“ auf, innerhalb dessen wir Objekte und Regeln des Fachgebietes darstellen können. Was im Raum enthalten ist, kann verwendet werden, was nicht enthalten ist, ist einer Repräsentation nicht zugänglich.

Die Semantik eines Fachgebiets kann umso präziser und widerspruchsfreier abgebildet werden, je mehr Achsen verfügbar, d. h. für den Anwender transparent und in Anwendungen einfach zu handhaben sind. In dieser Beziehung ist ein multifokales System einem mehrdimensionalen und einem Hie-

rarchiebaum entscheidend überlegen. Nicht zufällig verfügen Ontologien in der Regel über ein multifokales System.

Aus verständlichen Gründen wird für die Kommunikation (Interoperabilität) oft ein einfacheres System einem komplexeren vorgezogen. Nicht zufällig sind deshalb die gängigen Standards in der Medizin wie ICD-10 und CHOP typische Hierarchieebäume. Nachfolgend werden die damit verbundenen Probleme dargestellt.

5.5 Probleme bei der Abbildung einer komplexen Semantik in ein hierarchisches System

5.5.1 Nicht darstellbare Information

5.5.1.1 Prinzipielle Nichtausdrückbarkeit:

Die Abbildung 7 zeigt, wie die Information „Rezidiv/Nicht-Rezidiv“ in die ICD-10 GM eingefügt worden ist. In der ICD-10 WHO (Abbildung 6) fehlt diese Ausdrucksmöglichkeit. Die Information kann in diesem Standard bei Abdominalhernien prinzipiell nicht ausgedrückt werden. Das gleiche gilt, wie in Kapitel 4.4.1 beschrieben, auch für die Information „direkt/indirekt“. Nachfolgend stellen wir dar, dass solche fehlende Ausdrucksmöglichkeiten nicht durch ein schlechtes Design bedingt ist sondern ein prinzipielles Problem von baumartigen Ordnungssystemen darstellt.

5.5.1.2 Ausdrückbarkeit nur im Kontext

In Abbildung 6 und 7 finden sich zwischen den Ebenen der 3-Steller und der 4-Steller Kästchen für die Information „einseitig“ und „doppelseitig“; diese Information kann somit in der ICD-10 ausgedrückt werden. Allerdings nur mit Hilfe des 4-Stellers: Erst wenn die Information über die Komplikation ausgedrückt wird, wird die Information „einseitig/doppelseitig“ mit ausgedrückt. Andererseits verlangt der 4-Steller, dass man sich bei der Angabe der Komplikation auch zwingend über die „Einseitigkeit/Doppelseitigkeit“ ausdrückt. Diese enge formale Verknüpfung von zwei semantischen Freiheitsgraden stellt eine Einschränkung der Ausdrucksmöglichkeit dar, die in der Praxis umso schwerer wiegt, als bei einem Umgang mit dem Ordnungssystem dem Anwender die gegenseitigen Abhängigkeiten bewusst sein müssen, wenn er die Codierung korrekt durchführen will. Das gleiche gilt für das Lesen der so erhaltenen Codes. In der Praxis sind aber gerade diese Abhängigkeiten den Anwendern häufig nicht bewusst, was eine weitere Unschärfe der Codierung bewirkt. Diese Art Informationsproblem ist wie das vorhergehende prinzipiell in jedem Baum anzutreffen und lässt sich bei einer Baumarchitektur generell nicht vermeiden.

5.5.1.3 Informationsverlust durch Mehrfachbelegung

In den Abbildungen 6 und 7 findet sich die Information „einseitig/doppelseitig“ als Verzweigung im Hierarchiebaum und kann entsprechend ausgedrückt werden. Was aber codiert man, wenn diese Information nicht bekannt ist und man trotzdem einen 4-Steller ausgeben möchte oder muss? Die Systematik der ICD-10 verlangt in einem solchen Fall, dass die Codierung über „einseitig“ erfolgen soll und der entsprechende „einseitige“ 4-Steller auszugeben ist.

Was bedeutet das z.B. für den Code K41.4? – Es ist zwar klar und eindeutig, dass dieser Code eine

Femorale Hernie mit Gangrän bezeichnet, aber bezüglich „einseitig“ können wir nicht eindeutig schließen. Weil auch die Nichtkenntnis bezüglich „einseitig/doppelseitig“ mit diesem Code bezeichnet wird, kann K41.4 sowohl eine einseitige wie auch eine nicht näher bezeichnete, also möglicherweise doppelseitige Hernie bezeichnen. Die Information „einseitig“ kann somit durch keinen ICD-10 Code sicher repräsentiert werden.

5.5.2 Kombinatorische Explosion

Die kombinatorische Explosion ist in Abbildungen 6 und 7 erkennbar. Auf der Ebene der 4-Steller finden sich mehrere Dreier-Päckchen, die immer gleich aufgebaut sind. Dies rührt daher, dass die Information über die Komplikation (gelb) und die Information über den Ort der Hernie (grün) frei miteinander kombinierbar sind. Mathematisch handelt es sich um eine Multiplikation. Je mehr Merkmale (semantische Freiheitsgrade) durch ein Ordnungssystem abgebildet werden, umso mehr Faktoren fließen in die Multiplikation ein. Die Zahl der Kombinationen kann so sehr schnell explodieren.

Fehlende Ausdrückbarkeit (vgl. Kapitel 5.5.1) und kombinatorische Explosion (Kapitel 5.5.2) stehen in einem antagonistischen Zusammenhang: Je mehr versucht wird, alle Information darzustellen und keine Kombinationsmöglichkeit auszulassen, umso mehr gerät man in die kombinatorische Explosion. Um diese zu vermeiden, sind baumartige Ordnungssysteme immer „gestutzt“, d.h. es wird auf viele Kombinationsmöglichkeiten verzichtet. Dies aber führt genau wieder zu den Problemen von 5.5.1. Diese sind deshalb in nicht-multifokalen Systemen unvermeidbar.

Der Konstrukteur eines Baumes muss somit zwischen zwei Übeln auswählen: will er mehr nichtausdrückbare Information oder mehr kombinatorische Explosion? In der Praxis wird er versuchen zu optimieren, und versuchen, beide Probleme zu reduzieren. Dabei wird er aber unweigerlich von beiden etwas erhalten.

5.5.3 Verteilung einer singulären Information auf mehrere Codes

Nehmen wir an, jemand sucht alle Abdominalhernien mit Gangrän. Welche Codes muss er ansehen?

Aus Abbildung 7 wird klar, dass dafür acht 4-Steller und vier 5-Steller in Frage kommen. Eine einzelne Information (Komplikation: Gangrän) muss somit bei der ICD-10 an mehreren Stellen abgefragt werden.

Dabei handelt es sich um eine unvermeidliche Folge der kombinatorischen Explosion in einfachen baumartigen Ordnungssystemen. Die Explosion, so gedämpft sie auch sein mag, hat nicht nur Konsequenzen für den Umfang des Ordnungssystems sondern auch für die Weiterverwendung der Codes. Dadurch wird deren Verwendung für Auswertungen oder als Basis von Alert- und anderen Decision Support Systemen umständlich und fehleranfällig.

5.6 Probleme bei der Abbildung einer komplexen Semantik in ein multidimensionales System

Im Prinzip sind die im Vorkapitel genannten Probleme der hierarchischen Ordnungssysteme durch multidimensionale Systeme lösbar, indem jeder semantische Freiheitsgrad eine eigene Achse erhält. Dadurch werden die Werte auf den einzelnen Achsen beliebig miteinander kombinierbar und die Effekte der kombinatorischen Explosion treten nicht mehr auf.

Allerdings tauchen bei einem multidimensionalen System andere Probleme auf: Die Anzahl der Ach-

sen nimmt rasch zu und damit auch die Anzahl der Codes, da jede Achse über ihre eigenen Codes verfügen muss.

- a) **Achsenvermehrung:** Bei der konkreten Arbeit mit semantischen Systemen zeigt sich, dass die Achsenzahl schnell überhand nehmen kann. Allein für die Abdominalhernien können z.B. zehn unabhängige Freiheitsgrade gefunden und gefordert werden. Die Achsenvermehrung macht so ein System unhandlich. Wird aus diesem Grund auf gewisse Freiheitsgrade verzichtet, oder werden zwei oder mehr Freiheitsgrade in einer Achse zusammengefasst, ergeben sich die Probleme der baumartigen Systeme erneut (nicht ausdrückbare Information und/oder kombinatorische Explosion). Im Weiteren sind nicht alle Kombinationen sinnvoll. Z.B. kann man von einer offenen oder geschlossenen Fraktur sprechen, und offen/geschlossen ist somit ein Freiheitsgrad. Dieser macht neben den Frakturen auch für die Tuberkulose und andere Diagnosen Sinn. Für den Diabetes (offener Diabetes?) und die meisten Diagnosen macht dieser Freiheitsgrad aber überhaupt keinen Sinn. Mit zunehmenden Freiheitsgraden/Dimensionen in einem System wächst die Zahl sinnloser Kombinationen – bis praktisch alle Kombinationen sinnlos sind und das System fast nur noch leere Stellen enthält. Ein solches System ist weder übersichtlich noch performant. Die Probleme, die durch die Achsenvermehrung entstehen, können einfach dadurch gelöst werden, dass die Achsen gekapselt werden. Dadurch erhält man ein multifokales System (s Kapitel 5.3.7).
- b) **Codevermehrung:** Jede Dimension/Achse hat ihre eigenen Werte/Ausprägungen und wenn codiert wird, ihre eigenen Codes. Ein System mit drei Achsen braucht somit drei Codes für eine Diagnose. Dies war vor der Einführung von SNOMED CT das Problem von SNOMED: Es verlangte bei seiner Anwendung die separate Codierung der Morphologie, der Topologie und der Ätiologie – und gegebenenfalls von weiteren Freiheitsgraden. Aus dieser Kritik ist SNOMED CT entstanden. Jetzt fasst ein clinical term (CT) auf der präkoordinierten Aussenseite von SNOMED mehrere Codes des mehrdimensionalen, postkoordinierten Kerns zusammen. Der Anwender/Coder muss somit nur noch einmal, nämlich auf der präkoordinierten Aussenseite suchen. Allerdings entfallen durch dieses System mit den CT auch gewisse Vorteile der explizit mehrdimensionalen Repräsentation: Trotz des grossen Umfangs können auf der präkoordinierten Seite nicht alle Kombination des inneren Kerns repräsentiert werden, das System weist somit – wenn auch in stark abgeschwächter Form und nur auf der präkoordinierten Aussenseite – die in den Kapiteln 5.5.1 und 5.5.2 erwähnten Schwächen auf. Während die Nachteile der Achsenvermehrung durch ein multifokales Design überwunden werden können, bleibt das Problem der Vermehrung der Codes bestehen. Wie für multidimensionale muss auch für multifokale Systeme eine Entität (Diagnose) meistens mit mehreren Werten/Codes repräsentiert werden.

5.7 Mapping

5.7.1 Direktes Mapping

Die Herausforderungen des Mappings können exemplarisch an den oben aufgeführten Codierungen der Abdominalhernien gezeigt werden. Die Abbildungen 6 und 7 zeigen, welche Inhalte (semantische Freiheitsgrade) von den beiden untersuchten Codierwerken überhaupt dargestellt werden können. So kann z.B. die ICD-10 GM die Information, ob es sich bei der Inguinalhernie um ein Rezidiv handelt darstellen, die ICD-10 WHO aber nicht. Selbstverständlich geht bei einem Mapping der ICD-10 GM auf die ICD-10 WHO diese Information verloren. Umgekehrt, d.h. bei einem Mapping von ICD-10 WHO auf ICD-10 GM, kann die Information ebenfalls nicht verwendet werden, da sie primär (in der ICD-10 WHO) ja gar nicht auftaucht. Sie kann folgerichtig auch nicht in die Codes der ICD-10 GM übernommen werden kann. Egal, in welche Richtung das Mapping erfolgt, eine Information kann nur für Inhalte erfolgen, welche in beiden Codierwerken existieren.

Zur Beurteilung, welche Information zwischen zwei Codierwerken ausgetauscht werden kann, ist die semantische Analyse, wie in den Abbildung 6 und 7 dargestellt, unumgänglich. Dabei werden zuerst die Freiheitsgrade (Dimensionen), und anschliessend die Werte dieser Freiheitsgrade in beiden Systemen untersucht und verglichen.

Dabei gilt: Wenn in einem der Codierwerke der Freiheitsgrad fehlt, kann dieser Inhalt prinzipiell nicht gemappt werden. Wenn der Freiheitsgrad in beiden Codierwerken vorhanden ist, ist ein Mapping hingegen soweit – und nicht weiter – möglich, wie die Werte (Ausprägungen) des Freiheitsgrades in beiden Codierwerken aufeinander abbildbar sind.

Bei diesen Analysen für das Mapping muss nicht nur die prinzipielle Nichtausdrückbarkeit, sondern auch die Ausdrückbarkeit nur im Kontext und der Informationsverlust durch Mehrfachbelegung beachtet werden (siehe oben: 5.5.1.3).

Selbstverständlich ist es möglich, dass nicht beide Codierwerke gleich mächtig sind. In einem solchen Fall kann es leichter sein, die Information von einem Codierwerk A nach B abzubilden, als umgekehrt. Ganz abbildbar ist A auf B aber nur, wenn alle Freiheitsgrade und Werte von A auch in B darstellbar sind.

5.7.2 Mapping über eine Referenz

Während beim direkten Mapping zwei Codierwerke direkt verglichen werden, ist es auch möglich, das Mapping indirekt über eine möglichst umfassende Referenz durchzuführen. Abbildung 9 zeigt die Möglichkeiten des direkten Mapping von Codierwerk A und B in einer Mengendarstellung. Die Schnittmenge entspricht der Information, die in beiden Werken darstellbar und somit mappbar ist. Was sich ausserhalb der Schnittmenge befindet (blau und gelb gefärbt), entzieht sich dem Mapping.

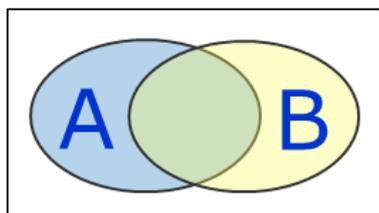


Abbildung 9: direktes Mapping

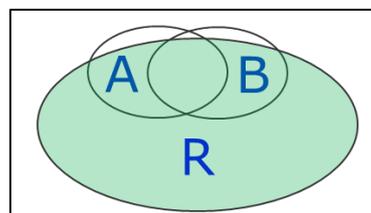


Abbildung 10: Mapping über eine Referenz R

In Abbildung 10 geschieht das Mapping von A und B über eine Referenz. Die Idee dabei ist, dass die Referenz dermassen umfassend ist, dass die Information von A wie auch von B in der Referenz darstellbar sind. Dabei besteht die Möglichkeit, dass trotz der gegenteiligen Absicht einige Informationen (Freiheitsgrade oder Werte) auch in der Referenz nicht darstellbar sind.

Wenn über eine Referenz gemappt wird, geschieht das eigentliche Mapping nicht direkt, sondern von A, bzw. B in die Referenz. Wie gross die Referenz auch ist, mappbar von A nach B ist natürlich nur, was in beiden Einzelsystemen darstellbar ist, also die gleiche Schnittmenge zwischen A und B wie in Abbildung 9 gezeigt.

Beim Mapping über eine Referenz also A und B jeweils gesondert und direkt in die Referenz gemappt. Die Referenz ist das eigentliche Zielsystem. Dazu muss es selber umfassend genug sein, um alle potentiell anbindbaren Systeme aufzufangen. Als weitere Aufgabe muss im Referenzsystem zeitnah die jeweilige Versionierung der einzubindenden Systeme eingepflegt werden.

5.7.3 Mapping über geführte Eingabe bzw. NLP (Mehrfachcodierung)

Eine weitere Möglichkeit besteht darin, dass gleich bei der Eingabe simultan in mehrere Systeme hinein codiert wird (=Mehrfachcodierung). Die semantische Interoperabilität wird so gar nicht mehr nötig, da alle relevanten Systeme bereits von Anfang an bedient werden. Es handelt sich in der Tat auch nicht mehr um ein Mapping, sondern um eine Alternative dazu.

Bei einer solchen Lösung stellen sich zwei prinzipielle Herausforderungen:

Erstens muss von Anfang an klar sein, welche Systeme bedient werden müssen – sonst muss für das neu dazukommende System retrospektiv neu codiert werden, oder dann muss doch ein Mapping zwischen den alten und dem neuen System mit allen oben genannten Problemen hinzugezogen werden.

Zweitens ist wohl kaum ein Anwender erfreut, wenn er die Aufgabe des ungeliebten Codierens gleich mehrfach zugewiesen bekommt. Deshalb wird eine Mehrfachcodierung in der Praxis nur dann realisierbar sein, wenn der Anwender in der Codierung so weit als möglich unterstützt wird. Für diese Unterstützung sind prinzipiell zwei Vorgehensweisen denkbar:

a) Geführte Eingabe über kaskadierte Menus:

Das Anwendersystem (KIS, RIS, etc.) führt den Anwender gezielt zur gewünschten Semantik, z.B. über ein Menu mit mehreren Stufen. Wenn es sich um relativ einfache Semantiken (z.B. in einem Labor) handelt, ist dies einfacher als bei komplexen, z.B. Diagnosen in der inneren Medizin. In letzterem Fall wird die Menukaskade schnell sehr ausgedehnt und die Eingabe kompliziert, fehleranfällig und ineffizient. Auch die Arbeit der Eingabe (Data Entry) wird für das medizinischen Personal mit den ausgedehnten Menukaskaden nicht einfacher, sondern aufwendiger.

Ein weiteres Problem ist, dass die Menus wirklich umfassend sein müssen. Und je umfassender sie sind, umso grösser ist die Zahl der Alternativen und umso länger dauert der Eingabeprozess. Für einfache und repetitive Eingaben können Kliniksysteme allerdings mit geführten Menus die Arbeit erleichtern.

b) NLP²⁵:

Alternativ zur geführten kaskadierten Menueingabe können NLP-Systeme natürlichsprachliche Eingaben im Hintergrund bearbeiten. Auf diese Weise können auch komplexe Semantiken „natürlich“ bedient werden. Solche NLP-Systeme müssen allerdings wirklich umfassend sein. Sie müssen nicht nur alle Freiheitsgrade und Werte intern repräsentieren können, sondern darüber hinaus über Regeln zur automatischen Interpretation der Texteingaben verfügen. Dabei geht es u.a. um Synonyme, Ambiguitäten, kontextabhängige Bedeutungen und Interpretation von komplexen Semantiken. Als drittes müssen diese Systeme codieren und brauchen somit Codierregeln, die wie die Sprachinterpretation automatisiert abgearbeitet werden müssen. Dabei darf das System nicht einfach mögliche Codierungen vorschlagen (analog zu einer Google-Suche, bei der mehrere, auch unzutreffende Hits akzeptiert werden), sondern muss wenn immer möglich direkt zur korrekten Codierung führen. Selbstverständlich erwarten Anwender von einem solchen System eine grosse Zuverlässigkeit (hohe Precision, hoher Recall). In Public Health Anwendungen (z. B. epidemiologische Anwendungen) und im Abrechnungswesen sind Fehler eines solchen Systems wohl eher tolerierbar als bei klinischen Anwendungen: Fehlcodierungen (z. B. eine fehlende Negation oder eine fälschliche Negation) schaden unmittelbar dem Patienten und dürften damit potentiell auch Haftpflichtansprüche nach sich ziehen.

Ein solches NLP-System hätte allerdings den Vorteil, dass es im Prinzip mit jedem KIS über relativ einfache Schnittstellen kommunizieren könnte – während die kaskadierten Eingaben als Teil der KIS-Oberfläche proprietär zu jedem KIS neu aufgebaut werden müssen.

²⁵ NLP = Natural Language Processing

5.7.4 Mapping (Zusammenfassung)

Die Aufgabe des Mappings wird generell stark unterschätzt. Ein Mapping ist – wie in 7.1. und 7.2. dargestellt ist, kaum wirklich umfassend und korrekt durchführbar. Erfahrungen mit dem Mapping von ICD-9 zu ICD-10 oder von CHOP zu OPS zeigen die Schwierigkeiten bei der Durchführung in der Praxis deutlich. Alternative Lösungen mit geführten Eingaben sind u.E. nur für einfache Dokumentationsaufgaben sinnvoll. Bei komplexen Semantiken (internistische Diagnosen) werden die Kaskaden zu mühsam und zu unsicher. Solche Lösungen sind eher für technische Informationen sinnvoll. NLP-Lösungen andererseits sind sehr anspruchsvoll, könnten aber – bei guter Fertigungsqualität – dem Anwender die Codierarbeit weitgehend abnehmen und dabei gleichzeitig für eine Mehrfachcodierung im Sinn von 7.3. sorgen.

6 Data Entry: Was sind die Anforderungen bezüglich Code-Eingabe beim Anwender? Unterschied zwischen technischen und klinischen Daten

6.1 Herausforderungen für das Data Entry:

Wenn ein semantischer Standard praktischen Nutzen bringen soll, so muss er auch korrekt eingegeben werden. Je komplexer ein Standard ist, umso fehleranfälliger wird die Eingabe. Nach unserer Ansicht liegt hier die grösste Hürde für die Einführung (Akzeptanz), die Durchführung (Praktikabilität und Aufwand) und den Nutzen von semantischen Standards.

Der Aufwand der Dateneingabe bestimmt auch massgeblich die Kosten. So sind mit der Einführung der ICD-10 in den Schweizer Krankenhäuser z.T. sehr grosse Codierabteilungen entstanden. Deren Kosten müssen in der Erfolgsrechnung berücksichtigt werden.

Folgende Faktoren beeinflussen die Qualität der Dateneingabe:

a) Umfang des Standards:

Je grösser der Standard ist, d.h. je mehr Werte er umfasst, umso grösser wird die Fehlerquote.

b) Komplexität des Standards:

Je komplexer das Fachgebiet ist, umso komplexer wird ein Standard, wenn er die Informationen des Fachgebietes umfassend repräsentieren soll. Dabei wird die Komplexität nicht durch unverständliche Expertenausdrücke gesteigert (diese haben meist sehr präzise Definitionen), viel schwieriger sind allgemeine Ausdrücke in einem komplexen Arbeitsgebiet, welche die unterschiedlichsten Anwendungen und Bezüge haben. Deshalb ist der eigentliche klinische Bereich wesentlich komplexer als z.B. Labor und Röntgen. Aus diesem Grund ist es einfacher, Standards in letzteren Gebieten einzuführen als auf einer Klinik für Inneren Medizin. Der Widerstand der Ärzte gegenüber Standards hat hier eine nachvollziehbare Quelle.

Ein Fehler wäre es, die Erfahrungen von technischen Standards einfach auf semantische zu extrapolieren. Durch ihre höhere Komplexität stellen semantische Standards eigene Herausforderungen. Obwohl die Komplexität eines Standards für die Dateneingabe eine Herausforderung ist, ist auch klar, dass ein Standard, der diese Komplexität verschleiert, wegen seinen Unklarheiten ebenfalls Aufwand bedeutet. Zudem sind die Resultate dieses vereinfachenden Standards weniger aussagekräftig und stärker fehlerbehaftet.

c) Nähe der eingebenden Person zur Information:

Je besser die dokumentierende Person das Fachgebiet und den Fall kennt, umso präziser wird ihre Eingabe vornehmen. So ist es für eine Codierfachkraft oft schwierig, einen Operationsbericht zu lesen und für CHOP zu codieren. Obwohl sie die Codierregeln gut kennt, muss sie auch Fachausdrücke der Chirurgen kennen. Zudem muss sie die Informationen zum Fall oft mühsam aus vielen Stellen zusammensuchen, dem behandelnden Arzt wären sie ohne Nachschauen bekannt.

d) Nähe der Ersteller des Standards zur Anwendung:

Pathologen können für Pathologen gute Standards bilden (SNOMED). Epidemiologen können für Epidemiologen gute Standards bilden (ICD-10-WHO). Für die Kostenberechnung in Kliniken, z.B. als Ausgang für DRGs, ist die ICD-10 der WHO jedoch primär nicht gedacht und musste deshalb zuerst angepasst werden (ICD-10-GM). Die Sprache, in der sich der Standard dem Eingebenden präsentiert, muss möglichst seiner gewohnten Sprache entsprechen. Dies vereinfacht die Eingabe und erhöht die Qualität der Eingabe.

6.2 Lösungsvarianten für das Data Entry:

In diesem Kapitel konzentrieren wir uns auf die Eingabe eines komplexen Codes, anhand des Beispiels von SNOMED CT

a) Eingabe von präkoordinierten klinischen Termen (CTs)

SNOMED CT verfügt über mehrere 100'000 unterschiedliche klinische Terme. Diese Terme können dem Eingebenden in entsprechenden KIS-Tools leicht auffindbar präsentiert werden. Allerdings ist dabei zu beachten, dass wegen der kombinatorischen Explosion die Liste der CTs von SNOMED weiterhin unvollständig ist.

b) Eingabe der postkoordinierten Codes von SNOMED

Diese Lösung erlaubt theoretisch eine vollständige und präzise Eingabe, ist für den Anwender aber vermutlich zu aufwendig.

c) Strukturierte Eingabe über die KIS-Software

Die Software des Anwenders steuert die Eingabe. Dabei wird eine fachspezifische Auswahl an Begriffen und Eingaben angeboten. Die Übersetzung der Eingaben in den semantischen Standard wird von der Software übernommen.

Eine solche Lösung ist umso praktikabler, je geschlossener das Fachgebiet ist. Die Gefahr besteht, dass der Anwender seinen Fall im Angebot nicht aufgeführt findet (wenn das System zu wenig ausgebaut ist) oder dass er die angebotenen Kaskaden zu mühsam findet (wenn das System extrem ausgebaut ist).

Ein weiterer heikler Punkt ist, dass man mit dieser Lösung vom Goodwill der Software-Hersteller abhängig ist.

d) Eingabe über NLP

Diese Lösung überwindet die Nachteile der oben stehenden Methoden. Der Anwender schreibt Klartext, ist somit in seinen Eingaben nicht auf präformiertes beschränkt. Die NLP-Software führt bei der Eingabe eine semantische Analyse der natürlichsprachlichen Eingaben durch und übersetzt kontrolliert in den Code des semantischen Standards. Gleichzeitig erlaubt die semantische Analyse das Überprüfen der Eingabe auf Vollständigkeit und fragt vom Eingebende gezielt das Fehlende ab. Dadurch bleibt der Anwender in der Formulierung frei und die Dokumentation ist vollständig und entspricht den Vorgaben des semantischen Standards. Wenn die NLP-Software regelbasiert ist, kann der Interpretationsvorgang genau verfolgt, gepflegt und verbessert werden. Zudem kann im Hintergrund eine semantische Repräsentation verwendet werden, die auch auf Details Bezug nimmt, die von SNOMED ignoriert werden. Durch die automatisierte semantische Strukturierung können die erhaltenen strukturierten Begriffe einzeln in eine Datenbank gegeben und später abgefragt werden (Semantic Data Repository).

Der Aufwand zur Erstellung eines solchen automatisch interpretierenden NLP-System ist allerdings beträchtlich. Es muss sowohl die natürliche Sprache verstehen, wie auch die Definitionen des semantischen Standards.

Zusammenfassend kann gesagt werden, dass das Data Entry sehr unterschiedliche Anforderungen voraussetzt, je nachdem ob es sich um technische oder klinische Daten handelt. Was bei technischen Daten einfach ist, kann bei klinischen Daten sehr schwierig werden. Der Erfolg auf der technischen Seite muss in der Klinik neu und anders erarbeitet werden. Deshalb ist es sicher sinnvoll, die Einführung eines semantischen Standards in einem einfachen Gebiet zu beginnen, solange man sich bewusst ist, dass im komplexeren Gebiet ganz neue Schwierigkeiten warten.

7 Interoperabilitäts-Niveaus

7.1 Technische Interoperabilität

Die technische Interoperabilität ist eine zwingende Voraussetzung für jede Art von Kommunikation von Systemen. Als Urform dafür kann wohl die elektrische Telegrafie betrachtet werden. Technische Interoperabilität hat somit bereits eine lange Geschichte; erstmals nachweislich erwähnt wurde sie bereits 1753²⁶. Heute können Computer untereinander Daten austauschen und sie sind damit technisch interoperabel.

7.2 Syntaktische Interoperabilität

Das European Telecommunication Standards Institute (ETSI) grenzt von der technischen die syntaktische Interoperabilität ab und beschreibt sie wie folgt: „... is usually associated with hardware/software components, systems and platforms that enable machine-to-machine communication to take place. This kind of interoperability is often centered on (communication) protocols and the infrastructure needed for those protocols to operate“^{27,28}.

7.3 Semantische Interoperabilität

Im eHealth Kontext wird gemeinhin von semantischer Interoperabilität von Systemen gesprochen: Ein zum Sender semantisch interoperables Empfängersystem soll eine Nachricht nicht nur verarbeiten (speichern, weiterleiten etc.) sondern auch deren Bedeutung erkennen muss.

Bemerkenswerterweise schreibt C.E. Shannon 1949 im Klassiker zur technischen Interoperabilität „The Mathematical Theory of Communication“ zum Thema der semantischen Interoperabilität: „Frequently the messages have meanings ... These semantic aspects of communication are irrelevant to the engineering problem.“

Auf der Ebene Maschine ist das „Erkennen des Wortinhaltes“ nur beschränkt möglich, z. B. im Sinne eines regelbasierten Verhaltens. Das eigentliche Verständnis der Bedeutung einer Information erfolgt beim menschlichen Benutzer und ist damit handlungsleitend.

Da der Begriff „semantische Interoperabilität“ bereits für die Ebene der Kommunikation zwischen Systemen besetzt ist; ordnen wir das menschliche Verständnis von Information der Prozess Interoperabilität zu. Zulässig erscheint uns diese Abgrenzung deshalb, weil das menschliche Verständnis handlungsleitend ist und Handlungen gemeinhin als Bestandteile von Prozessen verstanden werden. Aus unserer Sicht ist das menschliche Verständnis implizit Teil der Prozessinteroperabilität. Die explizite Erweiterung des Begriffs „Prozess Interoperabilität“ ermöglicht uns in der weiteren Diskussion eine klare modellhafte Abgrenzung.

²⁶ Kowalk W. Rechnernetze [Internet]. 2002 [updated 2002 Mar 10; cited 2014 Okt 19]. Available from: <http://einstein.informatik.uni-oldenburg.de/rechnernetze/elektris.htm>

²⁷ ETSI - European Telecommunications Standards Institute - (2006): Achieving Technical Interoperability – the ETSI Approach. ETSI White Paper No. 3. By Hans van der Veer (Lucent Technologies) and Anthony Wiles (ETSI), October 2006.

²⁸ ETSI spricht korrekterweise von „syntactical“. Zur Erinnerung: In diesem Dokument verwenden wir „Syntaktik“ und „Syntax“ gleichbedeutend.

Auf der Ebene der Maschine kann das „Erkennen der Wortinhalte“ zwar nicht erreicht werden, doch sind approximative Schritte möglich. Ein erster Meilenstein wird erreicht, wenn die Semantik formal repräsentiert werden kann – z.B. über die Darstellung der semantischen Achsen, des Achsengerüsts und der Werte auf den Achsen. In einem zweiten Schritt kann versucht werden, aus Freitexten automatisiert auf die entsprechenden Werte in diesen „semantischen Raum“ zu schliessen (NLP). Auf diese Weise ist ein Mapping zwischen verschiedenen Ontologien und anderen Ordnungssystemen - wenn auch nicht problemlos - denkbar.

7.4 Prozessinteroperabilität

Die Prozessinteroperabilität befasst sich mit der Integration der Systeme in den Arbeitsablauf. Erwähnt wird primär die Medienbruch-freie Verarbeitung ohne „händische“ Intervention.

Im vorliegenden Zusammenhang interessiert uns mehr, dass ein Prozess auch das menschliche Verständnis der Information beinhaltet; dieses Verständnis muss auf Senderseite weitgehend identisch sein mit demjenigen beim Empfänger.

7.5 Verschiedene Modelle der Interoperabilität

7.5.1 Modell der EHR Interoperability Work Group.

Das Modell der EHR Interoperability Work Group unterscheidet zwischen technischer Interoperabilität, semantischer Interoperabilität und Prozessinteroperabilität²⁹.

7.5.2 Levels of Conceptual Interoperability Model (LCIM)

LCIM kennt 6 Levels entsprechend Tabelle 1³⁰.

²⁹ EHR Interoperability Work Group. Coming To Terms - Scoping Interoperability for Health Care. 2007; Available from: <http://www.hln.com/assets/pdf/Coming-to-Terms-February-2007.pdf>

³⁰ Wang W, Tolk A, Wang W. The Levels of Conceptual Interoperability Model: Applying Systems Engineering Principles to M&S. In Wainer GA, Schaffer CA, McGraw RM, Chinn MJ. Proceedings of the 2009 Spring Simulation Multiconference, Spring-Sim 2009. San Diego: SCS/ACM; 2009. p. 375-384. Available at: <http://arxiv.org/pdf/0908.0191>

Tabelle 2: Descriptive Role of LCIM³¹

Levels	Description of Interoperability at this level
L6(Conceptual)	Interoperating systems at this level are completely aware of each others information, processes, contexts, and modelling assumptions.
L5(Dynamic)	Interoperating systems are able to re-orient information production and consumption based on understood changes to meaning, due to changing context as time increases.
L4(Pragmatic)	Interoperating systems will be aware of the context (system states and processes) and meaning of information being exchanged.
L3(Semantic)	Interoperating systems are exchanging a set of terms that they can semantically parse.
L2(Syntactic)	Have an agreed protocol to exchange the right forms of data in the right order, but the meaning of data elements is not established.
L1(Technical)	Have technical connection(s) and can exchange data between systems
L0(No)	NA

Für die aktuelle Diskussion von Interoperabilität und eHealth genügt im Prinzip das Modell der EHR Interoperability Work Group; zusätzliche grenzen wir noch die syntaktischen Interoperabilität ab.

³¹ Quelle: Wang W, Tolk A, Wang W. The Levels of Conceptual Interoperability Model: Applying Systems Engineering Principles to M&S. In Wainer GA, Schaffer CA, McGraw RM, Chinn MJ. Proceedings of the 2009 Spring Simulation Multiconference, SpringSim 2009. San Diego: SCS/ACM; 2009. p. 375-84.

8 Expertengespräche

8.1 Einleitung

Als Ausgangspunkt für die fachliche und politische Diskussion um die Einführung semantischer Standards in der Schweiz sollen der heutige Gebrauch semantischer Standards und die betroffenen Stakeholder in der Schweiz untersucht werden. Vorgesehen ist eine Erhebung mittels Umfrage. Umfrage und Auswertung sind nicht Gegenstand dieses Papiers; hingegen wird ein geeignetes Vorgehen entwickelt.

8.2 Methodik

8.2.1 Teilnehmer der Expertengespräche

Für die Entwicklung einer geeigneten Umfrage und die Definition potentieller Adressaten war ein Workshop mit folgenden Teilnehmern geplant:

Dieter Baumberger
André Assimacopoulos
Marc Oertle

Aus terminlichen Gründen war am Workshop vom 8. April 2015 Marc Oertle nicht anwesend, weshalb das Gespräch mit ihm am 16. Juni 2015 stattfand.

An den Besprechungen fand jeweils in einem erweiterten Kreis statt; Mitglieder der Expertengruppe Semantik bzw. von ehealth Suisse waren wie folgt anwesend:

8. April 2015:

Walter Fierz (Co-Leiter Expertengruppe Semantik eHealth Suisse)
Adrian Schmid
Johannes Gnaegi
Reinhold Sojer

Für das Protokoll danken wir Catherine Bugmann.

16. Juni 2015:

Christian Lovis (Co-Leiter Expertengruppe Semantik eHealth Suisse)
Walter Fierz (Co-Leiter Expertengruppe Semantik eHealth Suisse)
Sang-Il Kim

Für das Protokoll danken wir Stefan Wyss

Die Expertengespräche fanden unter Leitung der beiden Autoren statt.

8.2.2 Fragenkatalog für die Expertengespräche

Wir vertreten die Meinung, dass mit einer allfälligen Erhebung nicht nur der heutige Gebrauch semantischer Standards erfragt werden soll: Für das erste Expertengespräch erarbeiteten wir deshalb einen Entwurf für eine Erhebung, welche auch Fragen zu strategischen und operativen Zielen enthält, die mit der Einführung semantischer Standards erreicht werden sollen.

Der Entwurf umfasst ebenfalls mögliche Adressaten.

Für das zweite Expertengespräch arbeiteten wir die Ergebnisse des ersten Treffens ein; das Resultat bildete die Grundlage für das zweite Gespräch.

8.2.3 Diskussion und Rapportierung an die Expertengruppe Semantik eHealth Suisse

Am 8. April und am 22. Juni 2015 berichteten wir der Expertengruppe Semantik über die bisherigen Arbeiten und Ergebnisse. Vorschläge und Anregungen nahmen wir entgegen; sie sind in das vorliegende Dokument integriert.

8.3 Resultate

Nachfolgend stellen wir die Ergebnisse der Diskussionen zu den einzelnen Themen dar:

8.3.1 Zielsetzung

Im ersten Expertengespräch zeigte sich, dass noch kein Konsens darüber besteht, welche übergeordneten Ziele mit einer Erhebung erreicht werden sollen; ein operatives Ziel wurde im ersten Expertengespräch wie folgt definiert: Es sollen sowohl die Stellungnahmen von Verbänden, Institutionen etc. als auch die Meinung von Experten an der Front erfragt werden.

Im zweiten Expertengespräch wurden Ziele und damit Nutzen einer Umfrage intensiv diskutiert; ebenfalls zur Sprache kamen mögliche Risiken. Es wurde das Fehlen einer Semantikstrategie für eHealth in der Schweiz und daraus folgend auch das Fehlen einer Zielsetzung für eine Umfrage konstatiert.

Es kam die Frage auf, ob die Durchführung einer Erhebung zum jetzigen Zeitpunkt (2015) überhaupt sinnvoll sei.

Als Argumente pro Durchführung wurde eine Steigerung der Problem Awareness betreffend semantischer Fragen bei den Betroffenen genannt. Es wurde aber auch die Befürchtung geäußert, dass zur Zeit (2015) noch zu wenig Kenntnisse und Problembewusstsein bei den Akteuren im Gesundheitswesen vorhanden sind, so dass eine Umfrage zu wenig Verwertbarem oder sogar irreführenden Resultaten führen könnte.

Zur Frage, ob zum jetzigen Zeitpunkt eine Umfrage sinnvoll sei, wurde im zweiten Expertengespräch kein Konsens gefunden.

8.3.2 Adressaten der Umfrage / Vorgehen

Das Koordinationsorgan eHealth weist bereits im Mandatsauftrag explizit darauf hin, dass „neben Gruppierungen der öffentlichen Hand, auch Federführende Organisationen (FFO) und Fachverbände“ zu berücksichtigen sind.

Aus dem formulierten operativen Ziel betreffend Stellungnahmen (vgl. Kapitel 8.3.1) ergeben sich die Zielgruppen, bei denen eine Erhebung allenfalls durchgeführt werden soll wie folgt:

- Verbänden, Institutionen etc.
- Experten an der Front

Betreffend Vorgehen herrschte im ersten Expertengespräch folgender Konsens: Eine allfällige Erhebung soll mittels Fragebogen erfolgen. Das operative Vorgehen (Papierbasiert, Webbasiert etc.) bleibt offen.

Verbände / Institutionen sollen aufgefordert werden, Personen an der Front zu suchen, die den Fragebogen ausfüllen.

Aus den Expertengesprächen und den Diskussionen mit der Expertengruppe Semantik eHealth Suisse resultierte eine bereinigte Adressatenliste; sie findet sich im Anhang.

8.3.3 Fragenkatalog

Herauskristallisiert hat sich, dass ein exploratives Vorgehen mit offenen Fragen bevorzugt wird.

Aus den Expertengesprächen und den Diskussionen mit der Expertengruppe Semantik eHealth Suisse resultierte ein bereinigter Fragenkatalog; sie findet sich im Anhang. Alle Fragen konnten in den Gesprächen nicht ausdiskutiert werden.

Im zweiten Expertengespräch wurden die beiden Ansätze Source driven Strategy und Usage driven Strategy (Was sollen wir teilen?) einander gegenübergestellt.

Aus der ersten Strategie würde sich ableiten, dass man in einer allfälligen Erhebung die Akzeptanz einiger (weniger) semantischen Standards (z. B. SNOMED-CT, LOINC) erfragt.

Demgegenüber würde bei einer Usage driven Strategy primär erhoben, für welche Anwendungsfälle die Betroffenen semantische Standards benötigen.

8.3.4 Veränderte Erwartungen an das vorliegende Dokument

Zu Beginn der Arbeit waren die Zielsetzungen wie folgt:

- 1) Vorgehen zur Abklärung des heutigen Einsatzes und der Nutzung von semantische Standards
- 2) Vermittlung der Grundlagen für eine Diskussion semantischer Standards“ über Institutions- und Fachdomänen-Grenzen hinweg
- 3) Empfehlungen zum Einsatz von semantischen Standards

Im Verlauf des zweiten Expertengesprächs und in der darauf folgenden Diskussion im Rahmen des Reportings zuhanden der Expertengruppe eHealth Schweiz haben sich die Erwartungen an das vorliegende Dokument wie folgt verändert:

Die Strategielosigkeit betreffend Semantik wurde als zentrales Problem lokalisiert. Die Autoren wurden aufgefordert, sich zu einer möglichen Semantikstrategie eHealth zu äussern.

Des Weiteren wird eine Puffer-problematik thematisiert (s. Kapitel 10.4)

9 Sicht der Autoren

Auftragsgemäss stellen wir in diesem Kapitel unsere Überlegungen dar.

Unsere Argumentation und die daraus resultierenden Schlussfolgerungen folgen drei Ebenen, die wir wie folgt benennen:

1. Ziele
2. Fachlich semantische Ebene
3. Umfeld

Die Reihenfolge ergibt sich wie folgt: 1) definiert die übergeordneten Ziele und ist somit handlungsleitend. 2) zeigt die fachlichen Rahmenbedingungen und steckt damit die Grenzen des Machbaren ab. Aus 3) ergeben sich schliesslich Sachzwänge, die berücksichtigt werden müssen.

9.1 Ziele

Übergeordnetes Ziel ist die Prozessinteroperabilität zwischen allen Akteuren im Gesundheitswesen.

Im Umfeld von eHealth hat sich der Begriff der „semantischen Interoperabilität“ breit etabliert; wie wir bereits in der Diskussion des Anwendungsfälle (vgl. Kapitel 4.3) gesehen haben, greift die semantische Interoperabilität allerdings als Problemlösung zu kurz.

Die HL7 EHR Interoperability Work Group unterscheidet zwar zwischen technischer Interoperabilität, semantischer Interoperabilität und Prozessinteroperabilität³²; HL7 hat seine Standards jedoch explizit terminolgieneutral entwickelt und verzichtet somit darauf, die Interoperabilität konsequent bis zur Prozessinteroperabilität durchzusetzen.

Auch IHE bezieht in dieser Frage nicht eindeutig Stellung: Während z. B. das Radiologie Profil Consistent Presentation of Images (CPI)³³ mit seinem Bemühen nach durchgehend einheitlicher Darstellung von Bildern das Human Interfacing und damit Aspekte der Prozessinteroperabilität einschliesst, werden letztere in den meisten anderen Profiles explizit ausgeschlossen: sehr viele Profiles sind mit einer Maschine – Maschine Kommunikation gemäss Vorgabe erfüllt; die adäquate Repräsentation der Information wird nicht weiter thematisiert.

Die mangelnde Konsequenz bei der Behandlung von Fragen der Prozessinteroperabilität bzw. der Ausklammerung setzt sich fort in allen Diskussionen um eHealth. Im nächsten Kapitel begründen wir, warum die Diskussion semantischer Standards mit der expliziten Zielsetzung der Prozessinteroperabilität geführt werden muss.

³² EHR Interoperability Work Group. Coming To Terms - Scoping Interoperability for Health Care. 2007; Available from: <http://www.hln.com/assets/pdf/Coming-to-Terms-February-2007.pdf>

³³ Consistent Presentation of Images maintains the consistency of presentation for grayscale images and their presentation state information (including user and notations, shutters, flip/rotate, display area, and zoom). It also defines a standard contrast curve, the Grayscale Standard Display Function, against which different types of display and hardcopy output devices can be calibrated. Thus it supports hardcopy, softcopy and mixed environments. Quelle: http://wiki.ihe.net/index.php?title=Consistent_Presentation_of_Images

9.2 Fachlich semantische Ebene

Im ersten Satz des vorangehenden Kapitels sprechen wir bewusst von Akteuren: nicht sofort augenfällig ist, dass diese Akteure sowohl Systeme als auch Menschen umfassen. Wir gehen nicht näher auf die direkte Kommunikation zwischen Menschen ein sondern halten nur fest, dass eine Zwischenschaltung von IT-Systemen die Komplexität der Kommunikation zwischen Menschen erhöht: Im einfachsten Fall ergeben sich anstelle einer Schnittstelle Mensch – Mensch die vier Schnittstellen Mensch (Autor) – Maschine, Maschine – Maschine und Maschine – Mensch (Empfänger).

In den Kapiteln 7.3 und 7.4 haben wir das Verstehen der Bedeutung von Zeichen durch den Menschen modellhaft der Ebene der Prozessinteroperabilität zugeordnet. Daraus leiten wir ab, dass das Ziel der Prozessinteroperabilität nicht durch ein syntaktisch und semantisch definiertes (definierte Codes für definierte Begriffe) standardisiertes Austauschformat, also das Wire Format zwischen Maschine und Maschine, erreichen lässt.

Nehmen wir bei unserem Anwendungsfall mit dem Patienten Max Muster an, eines der beteiligten IT Systeme erlaube die Abbildung von „tachykardem Vorhofflimmern“ mit einem Qualifyer „paroxysmal“; das andere mit dem Qualifyer „rezidivierend“. Für das Austauschformat gehen wir davon aus, dass ein expliziter Code für „paroxysmales tachykardes Vorhofflimmern“ besteht.

Im vorliegenden Kontext sind die beiden Bezeichnungen „paroxysmal“ und „rezidivierend“ wohl überlappend, aber nicht wirklich deckungsgleich in der Bedeutung.

Dies wirft u. a. folgende Fragen auf:

- 1) In wieweit sind solche Abweichungen tolerierbar?
- 2) Wie sollen solche Abweichungen quantifiziert werden?
- 3) Wie lassen sich solche Abweichungen verhindern?

Selbstverständlich würden sich solche Abweichungen verhindern lassen, in dem genügend Begriffe zur Verfügung gestellt werden. Allerdings muss dabei der kombinatorischen Explosion vorgebeugt werden. Am besten geschieht dies mit einem mehrdimensionalen, im Idealfall multifokalen System. Ein solches System verlangt jedoch eine Codierung mehreren Codes, d.h. mit einem Code für jede Achse. Die Eingabe sollte den Anwender bei dieser Codierung unterstützen (Data Entry).

Mit der bisherigen eHealth Praxis hilft die obige Argumentation gemäss Kasten nur beschränkt weiter: Die Implementierungsleitfäden definieren primär Austauschformate und damit das Wire Format. Dass die betroffenen Systeme die Information intern und damit für den Benutzer korrekt abbilden ist dadurch noch nicht gegeben; deshalb stellen wird die Forderung nach Prozessinteroperabilität.

Aus der Forderung nach Prozessinteroperabilität geht hervor, dass die ganze Kette „Mensch (Autor) – Maschine, Maschine – Maschine und Maschine – Mensch (Empfänger)“ betrachtet werden muss. Semantische Standards müssen zusätzlich zur Festlegung des Austauschformates auch sicherstellen, dass die Repräsentation beim Empfänger der Eingabe beim Sender entspricht. Der Einfachheit halber lassen wir grafische, multimediale oder anderweitig komplexe Techniken bei Seite und beschränken uns auf die textuelle Repräsentation. Die Forderung müsste dann zumindest lauten:

a) Data Entry darf keine operative Hürde darstellen: das heisst der Anwender muss von der Aufgabe der mehrdimensionalen Codierung möglichst entlastet werden, z.B. durch NLP Methoden.

b) Der Display Text des IT Systems beim Empfänger muss für diesen diejenige Bedeutung haben, welche der Sender mit seiner Eingabe gemeint hat. Wenn Sender und Empfänger Systeme mit unterschiedlicher Sprache betreiben, kann dies bereits eine grössere Herausforderung sein.

Der Behandlungspfad der Patienten involviert verschiedene Fachgebiete und unterschiedliche Institutionen, das elektronische Patientendossier wird aus multiplen Knowledge Domains gespiesen; der Informationsaustausch ist Domain übergreifend.

Ein Domain übergreifender Informationsaustausch erfordert ein Domain übergreifendes Verständnis der verwendeten Begriffe; Die Zuordnung von Begriffen und Objekten des semantischen Dreiecks sollte in allen Domains identisch sein. Bezeichnungen / Codes müssen demgegenüber nicht identisch sein; zwischen Bezeichnungen bzw. Codes und Begriffen darf allerdings keine Ambiguität herrschen.

Es genügt nicht, die Syntaktik durch HL7 und IHE vorzugeben und als Referenzterminologie beispielsweise Snomed CT zu definieren, da so immer noch nicht definiert ist, ob ein bestimmter Sachverhalt im Einzelfall durch ActRelationships in CDA oder durch einen postkoordinierten Term von SNOME CT ausgedrückt wird³⁴.

9.3 Umfeld

Die Entwicklung von Austauschformaten erfordert immer ein fundiertes Domain Knowledge um die betreffende Knowledge Domain modellhaft abzubilden. Für eine solche Modellierung müssen die relevanten Inhalte bekannt sein, die Repräsentation (z. B. Display Texte) ist den gegebenen Usanzen (z. B. dem üblichem Sprachgebrauch) anzupassen, Fallstricke müssen antizipiert werden etc.

Aus obigem folgt - und die Praxis bestätigt dies - dass Austauschformate zweckmässigerweise domainspezifisch erstellt werden. Dies geht parallel mit den Hegemonialansprüchen einer Vielzahl von Stakeholder Gruppen (Fachgesellschaften, unterschiedliche Berufsgruppen etc.). Durch HL7 und IHE sind die Freiheitsgrade betreffend Syntax bereits reduziert; betreffend abzubildender Inhalte und deren Modellierung und somit der zu verwendenden Begriffe wird man über die bestehende Menge an Stakeholdern innerhalb nützlicher Zeit wohl kaum Konsens erreichen können.

Daraus folgern wir, dass einem zentralistischen top-down Approach nicht Erfolg beschieden sein wird.

³⁴ Allerdings ist das nicht der einzige Grund, dass solche Definitionen nicht ausreichen. Einfache Umsetzbarkeit in der Praxis spielt für die Akzeptanz und für das nötige Qualitätsniveau eine entscheidende Rolle.

10 Empfehlungen / Massnahmen für die Umsetzung

10.1 Ausgangspunkt

Unsere Empfehlungen gehen von folgendem aus:

1. Implementierungsleitfäden für Austauschformate werden auch zukünftig von unterschiedlichen Autoren erstellt werden und in unterschiedlichen Knowledge Domains beheimatet sein.
2. Die Stakeholder werden unterschiedliche Begriffe, Bezeichnungen und Codes verwenden.
3. Unverzichtbar ist ein identisches Verständnis der Begriffe bei den Endbenutzern und nicht nur auf Ebene des Austauschformates; identisches Verständnis der Begriffe bei den Endbenutzern haben wir auf der Ebene der Prozessinteroperabilität angesiedelt.

10.2 Eigenschaften von Implementierungsleitfäden

Im Sinne der Prozessinteroperabilität müssen Implementierungsleitfäden auch die Repräsentation der Information bei den Endbenutzern beinhalten. Dies umfasst im einfachsten Fall die Definition von Display Texten in den Sprachen Deutsch, Französisch, Italienisch und Englisch; falls weitere Sprachversionen vorgesehen sind, ist die Liste der berücksichtigten Sprachen zu ergänzen. Für andere Repräsentationen (grafisch, multimedial etc.) gilt die Aussage sinngemäss. Applikationen müssen zwingend die im Austauschformat spezifizierten Repräsentation abbilden können; User Interfaces dürfen die Bedeutung der Information nicht verändern, z. B. durch Weglassen von Teilen oder durch veränderten Kontext.

Ebenfalls in allen Sprachen ist die Definition der verwendeten Begriffe vorzuhalten.

Wichtig ist dabei folgendes: In den Userinterfaces müssen Bezeichnungen verwendet werden, die der allgemeinen Usanz im Fachgebiet entsprechen; das Aufzwingen eines bestimmten Wortgebrauchs wird schnell zu Widerstand führen. Dort, wo der Sprachgebrauch unscharf ist, wird man im Zug der Einführung von eHealth eine Präzisierung erarbeiten müssen; dies ist allerdings nichts Neues: auch bei der Erarbeitung von klinischen Guidelines ist dies immer wieder unumgänglich.

Für patientenbezogene Dokumente und den patientenbezogenen Informationsaustausch müssen die verwendeten Bezeichnungen auf definierte Begriffe und bestimmte dafür verwendete Codes abgebildet werden. Diese müssen aus einigen wenigen Referenzterminologien stammen; andernfalls steigt der Koordinationsaufwand ins Unermessliche.

Die Wahl von Referenzterminologien muss aufgrund des Typs ihres Ordnungssystems erfolgen: Es muss sichergestellt sein, dass einerseits der semantische Raum so aufgespannt wird, dass die Objekte der fachlichen Anwendungsdomäne genügend fein granuliert dargestellt werden können und andererseits die kombinatorische Explosion nicht übermässig zunimmt.

Wünschbar wäre nebst einer Beschränkung auf wenige Referenz als Quelle auch ein Referenzinformationsmodell. Auch wenn HL7 mit CDA ein solches Modell vorgibt, ist nicht alles definiert: Wie eine Negation dargestellt wird und inwieweit Beziehungen über die Strukturen von CDA (z. B. ActRelationship) oder über postkoordinierte Begriffe (z. B. aus SNOMED CT) ausgedrückt werden, ist nicht definiert.

Für ausgewählte Anwendungsfälle sollen deshalb Vorgaben gemacht werden.

Bei nativ bereits strukturierter Information (z. B. Laborresultate) ist eine strukturierte und codierte Ver-

arbeitung wesentlich einfacher zu realisieren als z. B. bei der Anamnese. Es empfiehlt sich deshalb vorerst, solche Anwendungsfälle umzusetzen.

10.3 Pflege von Austauschformaten und Referenzimplementa- tion

Die bisherigen Leitfäden für Austauschformate für eHealth in der Schweiz wurden im Rahmen zeitlich limitierter Mandate erarbeitet.

Nicht abgedeckt sind damit notwendige Anpassungen aufgrund von Veränderungen der Rahmenbedingungen, die Korrektur von Fehlern und an neue Anforderungen.

Für Austauschformate sind somit zusätzlich zur Entwicklung auch die Zuständigkeiten und die Ressourcen für eine kontinuierliche Begleitung / Weiterentwicklung zu definieren bzw. bereitzustellen.

Von den Betroffenen wird immer wieder die Frage nach der Anwendbarkeit in der Praxis gestellt. Die Feststellung geht über die vorliegende Thematik hinaus; wir machen sie trotzdem: Für Implementationsleitfäden für Austauschformate sollten unbedingt zeitnah die entsprechenden Referenzimplementationen erstellt werden; die dafür benötigten Mittel sind vorzusehen.

10.4 Puffer

Die standardisierte Interoperabilität von IT Systemen verfolgt auch das wirtschaftliche Ziel, dass Interfaces nicht von Fall zu Fall immer wieder neu entwickelt werden müssen; im internationalen Markt von IT Systemen des Gesundheitswesens folgt daraus unmittelbar ein Bedarf für internationale Standards.

Ein weiterer Grund ist die Aufwändigkeit der Entwicklung von Standards; auch daraus folgt die Notwendigkeit eines internationalen Ansatzes.

Sachzwänge werden es jedoch immer wieder notwendig machen, in Teilbereichen Anpassungen vorzunehmen und/oder zumindest mit zeitlichen Verschiebungen bei der Umsetzung leben zu müssen. Dies trifft sowohl für einzelne Domains untereinander als auch für die Beziehung zwischen eHealth in der Schweiz und im Ausland zu.

Daraus ergibt sich der Bedarf nach Puffern; nur mit Puffern können (ev. zeitlich limitierte oder auch dauerhafte) Abweichungen von Policies, Datenmodellen und Abweichungen in den verwendeten Begriffen und deren Codierung gemanagt werden.

Als technisches Mittel bietet sich eine Weiterentwicklung eines Terminologie Servers mit zusätzlichen Funktionalitäten an, welche über die Verwaltung von Relationen zwischen Referenzterminologien, Value Sets und Attribut Bindings sowie die Pflege der einzelnen Ordnungssysteme hinausgeht.

Grob umrissen müssten für die gewünschte Pufferfunktion sowohl Datenmodelle als auch Terminologien abgebildet werden können. Zwischen den Elementen verschiedener Datenmodelle, Terminologien und deren Versionen müssten unscharfe Relationen abgebildet werden können. Es müsste möglich sein, die Qualität solcher Relationen zu dokumentieren und die Veränderungen im zeitlichen Verlauf darzustellen³⁵. Ein solches System existiert nach unserem Wissen weltweit nicht.

³⁵ Die Beschreibung stützt sich auf Ausführungen von Ch. Lovis während der gemeinsamen Betreuung einer Semesterarbeit an der Berner Fachhochschule (Technology Center) durch ihn und J. Bleuer im Wintersemester 2014/15.

10.5 Umfrage

Wenn eine Umfrage durchgeführt wird kommt ausschliesslich ein Use Case driven Vorgehen in Betracht. Ein Source driven Ansatz ist schon nur deshalb wenig sinnvoll, weil Value Sets ihre Elemente häufig aus mehr als einer Source beziehen; HL7 sieht dies explizit in dieser Weise vor.

Eine Umfrage müsste somit primär die Anwendungsfälle erfragen; als willkommener Nebeneffekt würde möglicherweise die Awareness zu semantischen Fragen an der Front verbessert.

Wir enthalten uns einer Empfehlung für oder gegen eine Umfrage, weil uns eine Aufwand-Nutzen Abschätzung faktisch nicht möglich erscheint.

10.6 Schlussfolgerung

Im Verlauf des zweiten Expertengesprächs und in der darauf folgenden Diskussion im Rahmen des Reportings zuhanden der Expertengruppe eHealth Schweiz wurde das Fehlen einer Semantikstrategie als eines der Grundprobleme lokalisiert. Wir folgen allerdings Alain Rector in seiner Meinung, dass Terminologie und Datenmodelle nicht unabhängig voneinander betrachtet werden dürfen. Gerne würden wir deshalb anstelle von „Semantikstrategie“ von „Interoperabilitätsstrategie“ sprechen.

Diese Interoperabilitätsstrategie muss explizit das Ziel der Prozessinteroperabilität verfolgen.

Betreffend der in Kapitel 10.2 aufgeführten Eigenschaften von Implementationsleitfäden zu Austauschformaten empfehlen wir entsprechende Guidelines; als Herausgeber sehen wir das Koordinationsorgan eHealth Suisse. Die Aufgabe ist in keiner Weise trivial; die Lokalisation ausgewählter Probleme und die Erarbeitung entsprechender Vorgaben erfordert fundiertes semantisches Fachwissen.

Für die in 0 dargestellte Pflege der erstellten Austauschformate müssen die benötigten Ressourcen bereit gestellt werden und die Verantwortlichkeiten zugeordnet werden; die Federführung sehen wir z.B. bei eHealth Suisse.

Die Notwendigkeit, das Problem der Puffer zu lösen ist inhaltlich unbestritten; allerdings werden die einzelnen Aspekte nicht von allen gleich priorisiert.

Unabhängig von der Priorität des Problems kann ein erweiterter Terminologieserver in Betracht gezogen werden (wie in Kapitel 10.4 umrissen).

Ein solcher Terminologieserver kann bereits für die Grundaufgaben (Verwaltung von Referenzterminologien, Ableitung von Value Sets, Bindings zu Attributen etc.) eingesetzt werden und es erscheint sinnvoll, den späteren Ausbau zu einem erweiterten System (wie Kapitel 10.4 erwähnt) von Beginn an zu berücksichtigen. Die Autoren empfehlen, eine entsprechende Konzeptarbeit zu veranlassen. Für den Betrieb müsste landesweit eine einzige Stelle verantwortlich sein: diese würde somit auch die Inhalte der Referenzterminologien für das schweizerische Gesundheitswesen definieren. Eine föderalistische Alternative sehen wir nicht.

Die folgende Aussage bezieht sich auf die patientenbezogene Dokumentation und nicht auf Ordnungssysteme für DRG, Abrechnungen, Public Health und Statistik: Für den Laborbereich sind aktuell in der Schweiz LOINC und SNOMED CT daran, sich als Referenzterminologien zu etablieren; es gibt keinen Grund, dies in Frage zu stellen; SNOMED CT gilt auch für weitere Anwendungen als erste Wahl. Grundsätzlich unterstützen wir dies; allerdings sind noch Fragen betreffend Data Entry bei SNOMED CT zu studieren.

11 Anhang

11.1 Fragenkatalog (Vorschlag der Autoren ergänzt mit Vorschlägen aus dem ersten Expertengespräch und aus der Expertengruppe Semantik)

11.1.1 Einleitung

Nach dem ersten Expertengespräch bestand folgender Konsens:

- Es sollen sowohl die Stellungnahmen von Verbänden, Institutionen etc. als auch die Meinung von Experten an der Front erfragt werden.
- Es wird ein exploratives Vorgehen mit offenen Fragen bevorzugt.

11.1.2 Fragen zur aktuellen Nutzung von strukturiertem und codiertem Informationsaustausch

- Erfolgt in der von Ihnen vertretenen Organisation heute ein strukturierter und codierter Informationsaustausch?
- Verwenden Sie dafür eigene Ordnungssysteme (Welche und wofür ? Welchen Bereich / Anwendungsfall? Interne oder externe Kommunikation?)
- Verwenden Sie dafür einen semantischen Standard?
Wenn ja, welchen? Wenn nein, weshalb? Haben Sie einen wirtschaftlichen Nachteil durch die Einführung eines Standards befürchtet (-> Z. B. Verlust der Kundenbindung) ?
- Geben Sie bitte für jeden Anwendungsfall folgendes an:
 - Beschreibung des Anwendungsfalls (ausgetauschte Information)
 - Betroffene Fachgebiete
 - Beteiligte (Verwendung für die externe Kommunikation / Institutions-interne Verwendung), Kommunikationspartner?
 - Andere Anwender (namentliche Nennung der Organisationen, welche die gewählte Lösung auch anwenden oder dies planen)?
 - Ist die Lösung historisch gewachsen oder eine explizite Wahl? Begründung der Wahl.
 - Was wurde mit der Einführung geändert?
 - Welches Ziele wurden mit der Einführung verfolgt? Wurden diese erreicht?
 - Nutzen gegenüber der früheren Praxis?
 - Nachteile gegenüber der früheren Praxis?
 - Mängel im Vergleich zur Erwartung?
 - Unerwartete Benefits?
 - Aufwand für Pflege und Schulung (Größenordnung in CHF oder Personenzzeit)?
 - Aufwand für den Gebrauch (Größenordnung in CHF oder Personenzzeit)?
 - Aufwand für die Umstellung (Initialaufwand, Daueraufwand, Größenordnung in CHF oder Personenzzeit)?
 - Was wurde als Hindernis bei der Einführung wahrgenommen?
 - Verwenden Sie heute bereits ein Mapping zwischen verschiedenen Codes? Wenn ja zwischen welchen, und für welche Anwendungsfälle, wie erfolgt das Mapping?

11.1.3 Fragen zu den Erwartungen an einen zukünftige strukturierten und codierten Informationsaustausch

- Besteht in Ihrer Organisation ein Bedürfnis nach einem strukturierten und codierten Informationsaustausch?
- Anwendungsfälle?
- Geben Sie bitte für jeden Anwendungsfall folgendes an:
 - Beschreibung des Anwendungsfalls (auszutauschende Information)
 - Betroffene Fachgebiete
 - Beteiligte (Verwendung für die externe Kommunikation / Institutions-interne Verwendung), Kommunikationspartner?
 - Andere Anwender (namentliche Nennung der Organisationen, welche die geplante Lösung bereits anwenden oder dies planen)?
 - Begründung der vorgesehenen Lösung.
 - Was soll mit der Einführung geändert werden?
 - Welche Ziele sollen mit der Einführung erreicht werden?
 - Nutzen gegenüber der jetzigen Praxis?
 - Nachteile gegenüber der jetzigen Praxis?
 - Befürchtete Mängel im Vergleich zur Erwartung?
 - Mögliche zusätzliche Benefits?
 - Geschätzter Aufwand für Pflege und Schulung (Größenordnung in CHF oder Personenzzeit)?
 - Geschätzter Aufwand für den Gebrauch (Größenordnung in CHF oder Personenzzeit)?
 - Geschätzter Aufwand für die Umstellung (Initialaufwand, Daueraufwand, Größenordnung in CHF oder Personenzzeit)?
 - Was wird als Hindernis bei der Einführung wahrgenommen?
 - Ist ein Mapping zwischen verschiedenen Codes vorgesehen? Wenn ja zwischen welchen, und für welche Anwendungsfälle, wie erfolgt das Mapping?

11.1.4 Fragen zu den Anforderungen der Befragten an eine strukturierte und codierte Informationsverarbeitung.

Die nachfolgenden Fragen gehen davon aus, dass die Befragten in ihren Anwendungsfällen Codes aus einer oder mehreren definierten Referenzterminologien verwenden. Die Beschaffenheit der Referenzterminologie bestimmt, für welche Zwecke die Codes verwendet werden können. Für die Definition semantischer Standards und damit die Wahl von Referenzterminologien ist es deshalb wichtig, welche Anwendungsfälle damit abgedeckt werden sollten. Diese Frage darf sich nicht nur auf den Datenaustausch beschränken sondern muss auch die vorgelagerten (Dateneingabe/Codierung) und nachgelagerte Prozesse (z. B. Decision Support) einbeziehen.

Benennen Sie das(die) Fachgebiet(e), für welche Sie semantische Standards benötigen?

Benennen Sie für jedes Fachgebiet die vorgesehenen Anwendungsfälle:

- Ordering?
- Reporting?
- Interne Statistiken?
- Externe Statistiken/Public Health?
- Aggregation von Information (z. B. Zusammenfassung von Diagnosen unter Oberbegriffen)?
- Leistungserfassung?
- Planung, Art der Planungsaufgaben?
- Prozesssteuerung?

- Qualitycontrolling, Security?
- Reasoning, Decision Support, Case Based Reasoning?
- Konzeptbasierte Übersetzung zwischen verschiedenen Sprachen
- Weitere?
- Sollen ausschliesslich nativ bereits strukturierte Information (z. B. Laborresultate) codiert werden oder soll auch primär nicht strukturierte Information (z. B. Anamnese) codiert werden?

11.1.5 Weitere Fragen

- Um Nachfragen zu ermöglichen soll die Befragung nicht anonym erfolgen. Es sollen deshalb die Kontaktdaten erfragt werden.

Im Hinblick auf eine stratifizierte Auswertung soll erfragt werden:

- Grösse der jeweiligen Institution, Struktur (z. B. dezentrale Struktur bei Heimpflege Organisationen), Anzahl involvierte Systeme, betroffene Fachgebiete?
- Funktion der Befragten (Administrativ, IT, Medizinisches Personal, Pflegerisches Personal, Fachgebiet)?
- Für die Etablierung semantischer Standards wäre es sinnvoll, dass deren Anwender bekannt sind. Bei den zukünftigen Anwendern von semantischen Standards soll deshalb der Wunsch nach einer Registrierungsmöglichkeit erfragt werden.
- Unbedingt vorzusehen ist Raum für Bemerkungen der Befragten.

11.2 Adressaten für eine Umfrage (Vorschlag der Autoren ergänzt mit Vorschlägen aus dem ersten Expertengespräch und aus der Expertengruppe Semantik)

Anmerkung: Die einzelnen Organisationen können grundsätzlich auf dem offiziellen Weg kontaktiert werden; dort wo Ansprechpartner namentlich bekannt sind sie im Feld „Kontaktaufnahme über:“ aufgeführt.

Tabelle 3: Adressaten

	Organisation / Institution	Einbezug in eine Erhebung	Bemerkung
Ärzterschaft	FMH	Ja	Bitte an FMH um Weiterleitung an die Fachgesellschaften Bei Kontaktaufnahme mit Experten jeweils Info an J. Wagner
Zahnärzte	SSO		
Pathologen			
Forensik (Gerichtsmedizin)			
Militär	KSD		
	FMCH Foed. Med. Chir. (Helv.)	Ja	Chirurgen werden ansonsten über FMH erreicht
Apotheker / Medikamente	Verein der Amts- und Spitalapotheker (GSASA)	Ja	
	pharmaSuisse	Ja	
	Swissmedic	Ja	
	Documed (Compendium Schweiz)	Ja	

	Ofac (Rechenzentrum)	Ja	
Berufsorganisationen Gesundheitswesen	Schweizerischer Verband der Berufsorganisationen im Gesundheitswesen (SVBG)	Ja	
Pflegefachpersonen	SBK (Schweizer Berufsverband der Pflegefachfrauen und Pflegefachmänner)	Ja	
Fachleute für med. techn. Radiologie	SVMTRA/ASTRM (Schweizerische Vereinigung der Fachleute für med. techn. Radiologie)	Ja	
	Schweizerische Gesellschaft für Radiologie		
	HUG		
	Schweizerischer Hebammenverband	Ja	
Stationäre Versorgung (Spitäler)	H+ (Die Spitäler der Schweiz)	Ja	
Heime, soziale Institutionen	CURAVIVA (Dachverband Heime und soziale Institutionen)	Ja	Curaviva anschreiben und fragen, ob Sie weitere Experten kennen.
Heime, soziale Institutionen (inkl. spezialisierter Softwareanbieter/Dienstleister)	Provider	Ja	
Heimpflege (nonprofit)	Spitex Verband (nationale Dachverband der Schweizer Non-Profit-Spitex)	Ja	
	Spitex privée	Nein	
Heimpflege (profit)	senesuisse	Nein	
Ambulante Versorgung	Medswiss (Dachverband der Ärztenetze)	Ja	
	Institut für Hausarztmedizin		
	Einzelne Ärztenetzwerke anschreiben (Argomed, Eastcare, Medix)		
	Ärztelasse, Newindex (Datenpool Ärzte)		
	Santésuisse (Datenpool Versicherer)		
Labor	Verband der medizinischen Laboratorien (FAMH), Schweizerische Union für Labormedizin (SULM)	Ja	Industrie könnte mit der SULM abgedeckt werden. Es gibt einen Industrievertreter in der SULM. Sicherstellen, dass er den Fragebogen erhält.
Ziviles Rettungswesen			eHealth Suisse besorgt die Adressen
SRK	SRK		
Transplantationsmedizin	HLA-Bestimmungslabor Zürich und BAG Transplantationswesen		
Med. Informatik	SGMI	Ja	SGMI ist eine gute Quelle, um Einzelpersonen zu identifizieren.
Industrie	IG eHealth		
	Institut für Praxisinformatik (IPI)	Ja	
Behörden	Bundesamt für Gesundheit	Ja	
	KUV (Analyseliste)		
	SBFI sind im AAL-Bereich am standardisieren		
	OBSAN		
	Bundesamt für Statistik (BFS)	Ja	
	BASAN (Bundesamt für Sanität)	Ja	
	GDK (Schweizerische Konfe-	Ev.	Wird von eHealth Suisse mit Georg Schielke

	renz der Kant. Gesundheitsdirektoren und –direktorinnen)		(GDK) abgeklärt (s. Vorgehen mit PLG-Organisationen).
	Kantonale Gesundheitsdirektionen	Ja	Wird von „eHealth Suisse“ mit Georg Schielke (GDK) abgeklärt.
	Vereinigung Kantonsärzte (VKS)	Ja	
	Kantonsapotheker		
	Kantonale Präventionsstellen		
Kostenträger	Suva	Ja	
	Forum Datenaustausch	Ja	
	Vereinigung der Codier und CodiererInnen		
	Nationaler Verein für Qualitätsentwicklung in Spitälern und Kliniken (ANQ)	Ja	
	ASA/SVV (Schweizerischer Versicherungsverband)	Ja	
	Swiss Insurance Medicine	Ja	
	SGV/SSMC (Schweizerische Gesellschaft der Vertrauens- und Versicherungsärzte)	Ja	
	Militärversicherung	Nein	Wird über Suva erreicht
	Santesuisse	Ja	
Standardisierung	HL7 Benutzergruppe Schweiz	Ja	
	IHE Suisse	Ja	
	GS1	Ja	
	IPAG	Ja	
eHealthprojekte in der Schweiz	e-toile	Ja	
	eHealth VD	Ja	
	Infomed	Ja	
	Ponte Vecchio	Ja	
	Rete sanitaria	Ja	
Internat. Kontakte	epSOS / EXPAND	Ja	
	Weitere internationale Kontakte	Nein	Erfolgt über IHE und HL7 etc.
Public Health	Public Health Schweiz	Ja	
Prävention	EviPrev	Ja	
Prävention	Beratungsstelle für Unfallverhütung (bfu)	Ja	
Prävention	Stiftung Gesundheitsförderung Schweiz	Ja	
Universität / Fachhochschulen	Ausgewählte Experten	Ja	
Impfungen			
	SGMC		
	Ergotherapie		
	Physioswiss		
	Logopädie		
	ErnährungsberaterInnen		
	ChiropraktikerInnen		
	Pädiatrie		
	Palliativ Care		
	Krebsregister	Ja	
Veterinärmedizin		Ja	
Unispitäler	HUG	Ja	
	USZ	Ja	
	Inselspital	Ja	
	CHUV	Ja	
	Universitätsspital Basel	Ja	
	Alle Kantonsspitäler (St. Gal-		

	len / Aarau / Luzern / Solothurn / Neuenburg etc.)		
Patientenvertretung		Ja	eHealth Suisse fragt SPO als PLG-Organisation an (s. Vorgehen mit PLG Organisationen)
	HON		
Alle Mitglieder der Expertengruppe Semantik bzw. die Teilnehmer an den Expertengesprächen I und II			

11.3 Anmerkung zur Umfrage

Die Fragen und Adressaten für eine allfällige Erhebung wurden im ersten Expertengespräch und mit der Expertengruppe Semantik intensiv diskutiert. Als Resultat ist insbesondere der Fragenkatalog umfangreich geworden; dadurch besteht die Gefahr eines niedrigen Rücklaufes. Bei der Durchführung der Umfrage wird man nicht umhin kommen, Schwerpunkte zu setzen und Fragen wegzulassen.

11.4 Glossar

11.4.1 Begriff

Unter einem Begriff verstehen wird unsere Vorstellung von einem bestimmten Objekt, also das, was wir unter diesem Begriff verstehen. Begriffe werden mit Bezeichnungen bezeichnet und beziehen sich auf Objekte. Das semiotische Dreieck veranschaulicht den Zusammenhang zwischen Bezeichnungen, Begriffen und Objekten.

11.4.2 Bezeichnung

Eine Bezeichnung bezeichnet einen Begriff.

11.4.3 Binding

HL7 bezeichnet mit „Binding“ die Zuordnung eines Value Sets zu einem (oder mehreren) codierten Elementen eines Dokumentes oder einer Nachricht.

11.4.4 Code

Leiner et. al. setzen „Code“ mit „Notation“ gleich; diese definieren sie als „Zeichenfolge, die einen Begriff oder Begriffskombination repräsentiert“. Ein Code ist also ein Bezeichner eines Begriffs; im Modell des semiotischen Dreiecks werden „Bezeichnung“, „Benennung“, „Wort“, „Code“, „Bit“, „Term“, „Signifikant“ (und weitere, Bezeichnungen, die hier nicht genannt werden) gleichbedeutend verwendet³⁶.

³⁶ Leiner F et. al. Medizinische Dokumentation. Lehrbuch und Leitfaden. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. 3rd ed. Schattauer Stuttgart; 2015.

Leiner et. al. weisen darauf hin, dass eine Notation „...gegebenenfalls durch ihren Aufbau deren Stellung im systematischen Zusammenhang (eines Ordnungs- bzw. Begriffssystems) eindeutig abbildet“³⁷.

Es ist wichtig, dass letzteres keine obligatorische Eigenschaft eines Codes ist. Codes sind sinnvollerweise kurz und können aus Zahlen und/oder Buchstaben bestehen; in irgendeiner Weise sprechend müssen Codes nicht sein. Nicht sprechende Codes sind primär für die maschinelle Verarbeitung bestimmt; der Endbenutzer sollte nach Möglichkeit nicht mit Codes konfrontiert werden.

Ein Beispiel für einen Code ist der LOINC Code 31315-5; er bezeichnet das Stäbchen-Bakterium *Coxiella burnetii*.

Anmerkung: Coxiella burnetii ist ein gramnegatives und aerob lebendes, nur ca. 0,4 µm langes Stäbchen-Bakterium. Es ist der Erreger des Q-Fiebers. LOINC ist die Abkürzung von „Logical Observations Identifiers Names and Codes“ und benennt ein Codesystem für Untersuchungs- und Testergebnissen aus Labor und Klinik.

11.4.5 Code System

Der HL7 Version 3 Standard: Common Terminology Services stellt fest: „For the purposes of this specification, the terms vocabulary, terminology, and code system are used interchangeably“³⁸.

Anmerkung: Mit „Vocabulary“ wird hier etwas salopp ein „Controlled Vocabulary“ bezeichnet; dieser Hinweis ist wichtig, weil hier nicht der natürlich gewachsener Sprachgebrauch mit all seinen Ambiguitäten gemeint ist.

„Code System“ ist damit gleichbedeutend mit „Ordnungssystem“; die Bezeichnung „Terminologie“ verwenden wir hier jedoch anders als HL7 (s. 11.4.17).

11.4.6 Granularität

In einem hierarchischen Ordnungssystem enthalten die Ebenen, die nahe bei der Wurzel sind wenige und die weiter entfernten mehr Klassen. Je mehr Klassen pro Ebene vorhanden sind, umso feingranulärer ist die Unterscheidung, je weniger umso grobgranulärer.

Während eine feingranuläre Beschreibung detaillierter ist, ist eine grobgranuläre übersichtlicher.

11.4.7 Klassifikation

Leiner et. al. definieren „Klassifikation“ als „Ordnungssystem, das auf dem Prinzip der Klassenbildung beruht. In einer Klasse werden alle Begriffe zusammengefasst, zwischen denen nicht unterschieden werden soll“³⁹.

Geoffrey C. Bowker und Susan Leigh Star definieren „Klassifikation“ in ihrem Buch „SORTING THINGS OUT: CLASSIFICATION AND ITS CONSEQUENCES“ wie folgt (Auszug)⁴⁰: „A classification is a spatial, temporal or spatio-temporal segmentation of the world. A ‘classification sys-

³⁷ Leiner F et. al. Medizinische Dokumentation. Lehrbuch und Leitfaden. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. 3rd ed. Schattauer Stuttgart; 2015.

³⁸ HL7 Version 3 Standard: Common Terminology Services (CTS). Available at: http://www.hl7.org/implementation/standards/product_brief.cfm?product_id=10

³⁹ Leiner F et. al. Medizinische Dokumentation. Lehrbuch und Leitfaden. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. 3rd ed. Schattauer Stuttgart; 2015.

⁴⁰ Bowker GC, Leigh sTar S. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press, 1999.

tem' is a set of boxes (metaphorical or literal) into which things can be put in order to then do some kind of work - bureaucratic or knowledge production. In an abstract, ideal sense, a classification system exhibits the following properties: There are consistent, unique classificatory principles in operation..."

"The categories are mutually exclusive. In an ideal world, categories are clearly demarcated bins, into which any object addressed by the system will neatly and uniquely fit..."

"The system is complete. With respect to the items, actions or areas under its consideration, the ideal classification system provides total coverage of the world it describes. So, for example, a botanical classifier would not simply ignore a newly discovered plant, but would always strive to name it..."

Semantik und Metadaten Empfehlungen I⁴¹ definiert: „Eine Klassifikation beinhaltet die systematische Ordnung von Sachverhalten nach festgelegten Merkmalen, die mehrere Sachverhalte gemeinsam auszeichnen (Prinzip der Klassenbildung)“.

Anmerkung: Die obigen Definitionen sind kongruent; am anschaulichsten ist wohl diejenige von Geoffrey C. Bowker und Susan Leigh Star.

Klassifikationen sind also Einteilungen der Objekte eines bestimmten Gebietes. Dabei werden bestimmte Merkmale der Objekte berücksichtigt und andere ignoriert. Die Frage, welche berücksichtigt und welche ignoriert werden, enthält ein willkürliches Moment.

Die Klassen einer Klassifikation können als „Schubladen“ verstanden werden, in denen alle Objekte abgelegt werden, für welche die klassifikationstypischen Merkmale identisch sind.

11.4.8 Mapping

Wenn Werte aus einem Ordnungssystem in solche ein anderen abgebildet werden, spricht man von Mapping. Dies ist immer mit einem Informationsverlust verbunden, der unter Umständen erheblich sein kann.

11.4.9 Ontologie

Leiner et al. setzen „Ontologie“ der Bezeichnung „Begriffssystem“ gleich; letzteres definieren sie als „Menge von Begriffen, die entsprechend der zwischen ihnen bestehenden Beziehungen geordnet sind“⁴².

Semantik und Metadaten Empfehlungen I definiert: „Eine Ontologie bezeichnet eine sprachlich gefasste und formal geordnete Darstellung von Begriffen eines Anwendungsbereichs und der zwischen diesen Begriffen bestehenden Beziehungen“⁴³.

Anmerkung: Ontologien enthalten somit zusätzlich zu den hierarchischen Relationen, wie sie auch in hierarchischen Bäumen zwischen den Klassen vorkommen (is-a und is-part-of) weitere und unterschiedliche Qualitäten von Relationen. Diese bilden die Beziehung zwischen den Begriffen ab.

Diese Relationen sind zwar notwendige Elemente, jedoch keine Besonderheit der Ontologien, da sie bereits mit multifokalen Ordnungssystemen ausgedrückt werden können. Ontologien ermöglichen aber darüber hinaus das Ziehen von Schlüssen mit Hilfe dieser Relationen.

⁴¹ "eHealth Suisse". Semantik und Metadaten Empfehlungen I [Internet]. Bern: Koordinationsorgan eHealth Bund-Kantone; 2013 Jan 17 [cited 2013 Oct 9]. Available from: <http://www.e-health-suisse.ch/umsetzung/00146/00148/00238/index.html?lang=de>

⁴² Leiner F et al. Medizinische Dokumentation. Lehrbuch und Leitfaden. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. 3rd ed. Schattauer Stuttgart; 2015.

⁴³ "eHealth Suisse". Semantik und Metadaten Empfehlungen I [Internet]. Bern: Koordinationsorgan eHealth Bund-Kantone; 2013 Jan 17 [cited 2013 Oct 9]. Available from: <http://www.e-health-suisse.ch/umsetzung/00146/00148/00238/index.html?lang=de>

Was Ontologien (so wie wir die Bezeichnung hier verwenden) somit kennzeichnet, ist ein logisches Regelwerk und somit die Fähigkeit zwischen verschiedenen Freiheitsgraden Schlüsse zu ziehen

11.4.10 Ordnungssystem

Leiner et al. definieren „Ordnungssystem“ wie folgt „...Dokumentationssprache, die auf einer Begriffsordnung basiert. Es definiert und beschreibt alle Deskriptoren oder zugelassenen Benennungen und ordnet sie entsprechend der Ordnung der Begriffe“⁴⁴.

Ein „Ordnungssystem“ ist nach Leiner et al. eine Dokumentationssprache, also eine formalisierte Sprache mit einer definierten Grammatik.

In diesem Dokument verwenden wir die Bezeichnung „Ordnungssystem“ anders, nämlich als Oberbegriff aller Systeme der Benennung und/oder Ordnung.

11.4.11 Präkoordination / Postkoordination

Die beiden Ausdrücke werden von SNOMED verwendet. SNOMED bezeichnet die ausgedehnte Auflistung von klinischen Termen (clinical terms = CT) als präkoordiniert und Begriffe des mehrdimensionalen, strukturierten Ordnungssystem, auf die es diese ungeordneten präkoordinierten Terme abbildet, als postkoordiniert.

11.4.12 Prozessinteroperabilität

Die Prozessinteroperabilität befasst sich mit der Integration der Systeme in den Arbeitsablauf. Im Zentrum steht die Medienbruch-freie Verarbeitung ohne „händische“ Intervention. Nebst grösserer Effizienz ist vor allem das Vermeiden von Fehlern (z. B. beim Abschreiben von Laborresultaten) ein wichtiger Nutzen von prozessinteroperablen Systemen.

11.4.13 Referenzterminologie

Genau genommen müsste es „Referenz-Ordnungssystem heissen, da wir in diesem Dokument für den Oberbegriff aller Systeme der Benennung und/oder Ordnung die Bezeichnung „Ordnungssystem“ verwenden. Die Bezeichnung „Referenz-Ordnungssystem“ ist allerdings unüblich; wir verwenden deshalb auch „Referenzterminologie“ und bezeichnen damit eines (von möglichst wenigen) Code Systems (Ordnungssystemen), aus denen die Value Sets für codierte Elemente abgeleitet werden.

11.4.14 Semantik

Semantik beschäftigt sich mit der Bedeutung von Zeichen, Daten, Wörtern usw.

In Semantik und Metadaten Empfehlungen I wird erklärt: „Die „Semantik“ oder „Bedeutungslehre“ befasst sich mit der Bedeutung von Zeichen. „Zeichen“ können dabei Wörter, Ausdrücke oder Symbole sein. Ein Wort oder Ausdruck kann je nach Kontext etwas anderes aussagen. So kann das Wort

⁴⁴ Leiner F et. al. Medizinische Dokumentation. Lehrbuch und Leitfaden. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. 3rd ed. Schattauer Stuttgart; 2015.

„Bank“ ein Geldinstitut oder ein Ort zum Ausruhen bedeuten. Erst wenn die Bedeutung einer Information klar ist, wird eine nahtlose Kommunikation zwischen heterogenen technischen Informationssystemen für alle Verwendungen und alle beteiligten Akteure möglich.“⁴⁵.

11.4.15 Semantische Interoperabilität

Ein zum Sender semantisch interoperables Empfängersystem muss eine Nachricht nicht nur verarbeiten (speichern, weiterleiten etc.) sondern auch deren Bedeutung erkennen. Der Begriff „Bedeutung“ ist dabei in seiner einfachsten Form zu verstehen im Sinne von „Erkennen des Wortinhaltes“. Konkret geht es um die Vermeidung von Kommunikationsproblemen aufgrund von Mehrdeutigkeit.

11.4.16 Semiotisches Dreieck

Das semiotische Dreieck veranschaulicht die Beziehungen von Bezeichnungen, Begriffen und Objekten untereinander.

Es gelten folgende Beziehungen: Eine Bezeichnung bezeichnet (oder benennt) einen Begriff, ein Begriff bezieht sich auf eine Klasse von Objekten. Dadurch bedeutet eine Bezeichnung eine Klasse von Objekten.

Zur Illustration sei eine Parallele gezogen zur Objektorientierten Programmierung (OOPP: So wie wir in der OOP zwischen Objekttypen (Klassen) und Instanzen unterscheiden, unterscheiden wir in der Semantik zwischen Begriffen und den Objekten⁴⁶, auf welche die Begriffe hinweisen⁴⁷.

Das Semiotische Dreieck hat eine lange Geschichte: Bereits Platon hat ein Wort-Gegenstand-Modell zwischen Namen (Zeichen), Idee (Begriff) und Ding verwendet.

11.4.17 Terminologie

Leiner et. al. definiert „Terminologie“ als „geordnete Zusammenstellung der Benennungen eines Fachgebietes, zusammen mit Definitionen der Begriffe, welche sie benennen“⁴⁸.

In Semantik und Metadaten Empfehlungen I finden sich folgende Definition: Eine Terminologie umfasst definierte Begriffe aus einem bestimmten Fachgebiet⁴⁹.

In ISO 1087 ist „terminology“ definiert als „Set of terms representing the system of concepts of a particular subject field“⁵⁰.

DIN 2342 definiert „Terminologie“ wie folgt: „Gesamtbestand der Begriffe und ihrer Benennungen in

⁴⁵ "eHealth Suisse". Semantik und Metadaten Empfehlungen I [Internet]. Bern: Koordinationsorgan eHealth Bund-Kantone; 2013 Jan 17 [cited 2013 Oct 9]. Available from: <http://www.e-health-suisse.ch/umsetzung/00146/00148/00238/index.html?lang=de>

⁴⁶ Die Objekte müssen nicht zwingend real sein, sondern können auch Phantasieobjekte sein (z. B. ein Einhorn, eine Expedition zur Andromeda). Im Einhornbeispiel gibt es wieder beides: den Begriff „Einhorn“ und das damit gemeinte Objekt, nämlich das konkrete Einhorn in meiner Phantasie. Im semiotischen Dreieck finden sich die Objekte immer an der rechten Ecke.

⁴⁷ Zur Veranschaulichung des Unterschieds sei als Beispiel das Stehpult von J.W. von Goethe erwähnt. „Stehpult“ ist ein Unterbegriff von „Pult“, „Pult“ ist ein Unterbegriff von „Möbel“. Alle drei Begriffe sind Begriffe in unserem Kopf. Das konkrete Stehpult aber (die reale Instanz) kann sich unmöglich in unserem Kopf befinden.

⁴⁸ Leiner F et. al. Medizinische Dokumentation. Lehrbuch und Leitfaden. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. 3rd ed. Schattauer Stuttgart; 2015.

⁴⁹ "eHealth Suisse". Semantik und Metadaten Empfehlungen I [Internet]. Bern: Koordinationsorgan eHealth Bund-Kantone; 2013 Jan 17 [cited 2013 Oct 9]. Available from: <http://www.e-health-suisse.ch/umsetzung/00146/00148/00238/index.html?lang=de>

⁵⁰ISO 1087-1:2000. Available at: http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=20057

einem Fachgebiet.“⁵¹.

Anmerkung: Die Definition von Leiner et. a. fordert im Gegensatz zu den anderen Definitionen eine „Geordnetheit“⁵². Dies erscheint und nicht nötig, zumal diese Forderung auch nicht der Sichtweise von HL7 entspricht. In der Definition von Semantik und Metadaten Empfehlungen I ist nicht erwähnt, dass eine Terminologie nicht nur die Begriffe, sondern auch ihre Benennungen enthält. ISO und DIN sind im Wesentlichen gleich; ISO erscheint uns etwas präziser, weshalb wir uns an die ISO Definition halten.

Gleich wie ISO 1087 unterscheidet auch DIN 2342 zwischen „Benennung“ und „Begriff“; die Definition von ISO 1087 ist jedoch präziser; wir halten uns deshalb an diese und verwenden die Bezeichnung „Terminologie“ für Ordnungssysteme mit Betonung der Bezeichnung und ohne explizite innere Struktur.

11.4.18 Value Set

HL7 definiert „Value Set als „...uniquely identifiable set of valid concept representations, where any concept representation can be tested to determine whether or not it is a member of the value set“⁵³.

⁵¹ DIN 2342. Begriffe der Terminologielehre. Available at: <http://www.nat.din.de/cmd?level=tpl-art-detailansicht&committeeid=54739043&artid=140370128&languageid=de&bcrumblevel=1&subcommitteeid=54749613>

⁵² Leiner F et. al. Medizinische Dokumentation. Lehrbuch und Leitfaden. Grundlagen einer qualitätsgesicherten integrierten Krankenversorgung. 3rd ed. Schattauer Stuttgart; 2015.

⁵³ Quelle: http://wiki.hl7.org/index.php?title=Domain_and_Value_Set_Definitions_and_Binding#Domain_and_Value_Set_Definitions_and_Binding

Ein Value Set nach HL7 definiert damit die Menge der Konzepte mit ihren Bezeichnungen (Codes), welche für ein codiertes Element eines Dokumentes oder einer Nachricht verwendet werden dürfen. Ein einzelnes Value Set kann Elemente aus verschiedenen Code Systems (Ordnungssystemen) enthalten.

11.5 Synopsis der semantischen Bezeichnungen und Begriffe

Semantische Beschreibung	Dafür häufig verwendete Bezeichnungen	In diesem Dokument verwendete Bezeichnungen	Von eHealth Suisse verwendete Bezeichnung	Von HL7 verwendete Bezeichnung
Entitäten auf der Ebene der Kommunikation	Ausdruck, Bezeichnung, Term, Wort	Bezeichnung	Bezeichnung	Code
Entitäten der semantischen Ebene	Begriff, Konzept	Begriff, Konzept	Begriff, Konzept	Concept
System der Benennung und/oder Ordnung (= Oberbegriff für alle weiteren, unten genannten Systeme)	Ordnungssystem, Terminologie, Klassifikationssystem, Codesystem	Ordnungssystem	Ordnungssystem	Code System
- mit Betonung auf der Benennung	Nomenklatur, (kontrolliertes) Vokabular, Terminologie, Thesaurus	Terminologie	Nomenklatur	
- mit Betonung auf der Ordnung	Klassifikation	Klassifikation	Klassifikation, Ontologie	
Eindimensionales Ordnungssystem	Klassifikation, Hierarchische Klassifikation, Taxonomie	Hierarchisches Ordnungssystem	Klassifikation	

Mehrdimensionales Ordnungssystem	Klassifikation, mehrdimensionale Klassifikation	Mehrdimensionales Ordnungssystem	Klassifikation	
Multifokales Ordnungssystem	Klassifikation, multifokale Klassifikation	Multifokales Ordnungssystem	Klassifikation	
Ordnungssysteme für Ontologen (Begriffe und Relationen mit logischem Regelwerk)	Ontologie	Ontologie	Ontologie	

Beispiele				
ICD-10, CHOP		Hierarchisches Ordnungssystem	Klassifikation	Code System
SNOMED CT, ICNP		Mehrdimensionales Ordnungssystem	Klassifikation	Code System

11.6 Synopsis der Ordnungssysteme

