

## Codierung als Interpretationsvorgang

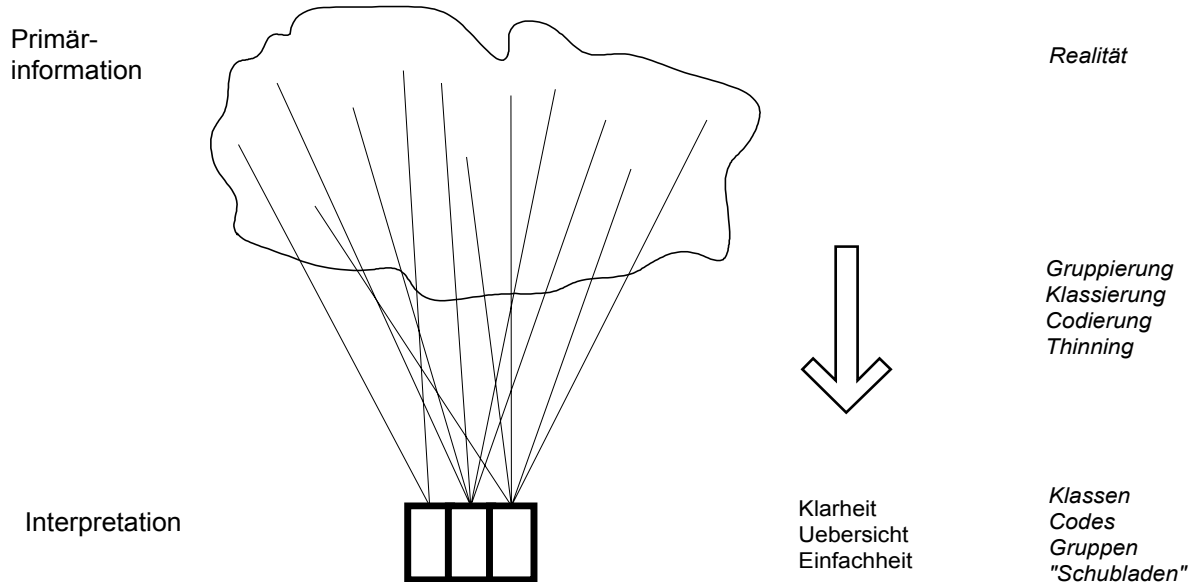


Abb. 1: Der Codiervorgang

Bei jeder medizinischen Codierung wird ein Sachverhalt vereinfacht. Diese Vereinfachung kann als eine Interpretation der zugrundeliegenden medizinischen Daten angesehen werden - genauso wie die medizinischen Daten ihrerseits eine Vereinfachung und Interpretation des viel komplexeren, real existierenden Krankheitsfalles sind.

Dabei ist evident:

1. Die Primärdaten, welche sich immer auf die konkreten Einzelfälle beziehen, sind reicher strukturiert und tragen mehr Information in sich als später für eine Auswertung oder Interpretation gewünscht wird.
2. Der Prozess der Interpretation beinhaltet eine Verarmung, nämlich eine Reduktion der Anzahl der Daten (=bits). Nur durch die vereinfachende Strukturierung der Daten kann eine verwertbare Aussage gewonnen werden und dies bedeutet eine Reduktion der ursprünglichen Gesamtinformation.
3. Welche Information wird weggelassen? Die Frage ist keinesfalls banal, denn die Auswahl entscheidet, welche Information schliesslich übrig bleibt. Man könnte nun annehmen, dass es immer eine quasi naturgegebene Auswahl gibt. Dies ist nicht der Fall. Denn die Auswahl liegt nicht nur in den betrachteten Daten - dem "Objekt" - begründet, sondern auch in der Absicht der Untersuchung, d.h. im Kontext der Fragestellung - im "Subjekt" also. Je nach Fragestellung ändert die Sichtweise und somit auch die Auswahl der für sie relevanten Information.
4. Um bei diesem Weglassen unter grösstmöglicher Uebersichtlichkeit möglichst viele Details und Sichtweisen auf der interpretierenden Seite zu erhalten, kommt der Strukturierung der Information

auf der Interpretationsseite und somit dem Modell der Wissensrepräsentation (Begriffsarchitektur) eine hohe Bedeutung zu. Architekturen, welche die Komplexität der Ausgangssituation besser einfangen, können mit weniger Daten (bits) mehr Informationen (Aussagen) repräsentieren und so die Primärdaten klarer wiedergeben (interpretieren).

Zu diesen vier Punkten möchte ich folgendes anmerken:

ad 1:

Der Ausdruck "*Primärdaten*" meint die am Ort ihrer Entstehung gewonnenen Daten, welche möglichst umfassend der Vielfalt des wirklichen Lebens entsprechen, Es gibt nichts, was mehr Detailinformation enthält als die Realität. Andererseits ist eine bewusste Entscheidung erst bei einer klaren Sicht der Dinge möglich. Bei zuvielen Bäumen ist der Wald nicht mehr erkennbar. Die klare Sicht darf nicht allzu detailreich sein, sondern muss einfach sein und zuvorderst das Wesentliche enthalten. Im Gesundheitswesen ist mit den Primärdaten die Summe aller Informationen gemeint, die Ärzten und Pflegepersonal zugänglich sind, eine Unmenge an Daten also, die als Summe natürlich völlig hypothetisch ist, da sie in Wirklichkeit nie an einem Ort zusammengetragen werden kann. Schon beim Sammeln an der vordersten Front findet deshalb im einzelnen Beobachter - Krankenschwester, Arzt usw. - eine datenselektierende Interpretation der Art statt, wie sie hier erläutert wird. Das mit diesem Text thematisierte Phänomen der Datenselektion ist überall präsent, wo mit Information umgegangen wird.

ad 2:

Die Aussage der These 2 beinhaltet ein *Paradox*: Da die Gesamtmenge der vielen Details für uns noch keine klare Aussage enthält, vereinfachen wir die Menge der Informationen, bis wir aus der reduzierten Menge die Information gewinnen, die uns interessiert. Die resultierende Information (verwertbare Aussage) wird also gewonnen, indem die vorbestehende Information (Summe der Primärdaten) reduziert wird.

Der beschriebene Prozess der Informationsauswahl findet nicht nur im Speziellen zur Gewinnung von Statistiken und ähnlichem statt. Er geschieht viel allgemeiner überall dort, wo eine vielfältige Aussenwelt (eine Realität) interpretiert wird. Er geschieht also nicht nur in EDV-Systemen sondern auch in biologischen datenverarbeitenden Systemen (=BDV's). Die menschliche Netzhaut z.B. besteht aus ca. 125 Millionen einzelnen optischen Rezeptoren (=Pixeln). Eine entsprechende Menge an Signalen muss verarbeitet und ihre Gesamtdatenmenge muss dabei radikal redimensioniert werden, um schliesslich dem Bewusstsein eine sinnvolle Information präsentieren zu können.

Der Informationsfluss von der realen Vielfalt zur aussageträchtigen Vereinfachung ist stets auch ein *Fluss von der Peripherie zum Zentrum*. In der Biologie fließen die Information von den Sinneszellen der peripheren Organe (Haut, Netzhaut, Ohr usw) zum zentralen Organ (Hirn). Dort werden sie weiter vereinfacht bis sie als einfache Wahrnehmungen im (zentralen) Bewusstsein erscheinen. Aus Millionen von Pixeln auf der Netzhaut und vielen in der Erinnerung gespeicherten Mustern erscheint als Quintessenz z.B. die einfache Wahrnehmung: Hier ist ein Apfel. Analog fließen in einem Unternehmen die vielfältigsten Information von der Peripherie zum Zentrum, werden dort interpretiert und bilden als möglichst simple und stringente Geschäfts-"Wahrnehmungen" die Grundlage für die Handlungsentscheidungen. Der Fluss von der Peripherie zum Zentrum ist stets gekoppelt mit dem genannten *Informationsparadox der Dateninterpretation*.

Ad 3:

Weshalb wird so oft angenommen, dass es eine naturgegebene Vereinfachung der Welt gebe? Weshalb möchten wir gerne an eine "objektive", alleingültige Klassifikation glauben? Der Grund ist der,

dass wir uns an den falschen Beispielen orientieren. Wir setzen in der Natur einen solchen hierarchischen, geschlossenen Bauplan voraus, wie er sich immer dann in den Konstrukten unseres Geistes findet, wenn wir uns Mühe geben, systematisch vorzugehen. Bankkonten, Archive, Lager, Maschinen, Computer und Verwaltungsorganigramme sind Konstrukte des menschlichen Geistes und zeigen in ihrer Anlage ein systematisches, oft hierarchisches Muster. Auf sie ist deshalb ein vergleichsweise einfaches Repräsentationssystem anwendbar.

Computeranwendungen in diesen technischen-administrativen Bereichen sind deshalb schneller von Erfolg gekrönt als Anwendungen zur Darstellung biologischer Variabilität. Eine Beobachtung, die gerade im Klinikbereich höchst augenfällig ist, wo die Administration um Jahrzehnte vor dem ärztlichen Bereich computerisiert wurde. Der Erfolg von einfachen Einteilungsschemen in technischen Bereichen verleitet uns, einfache Modelle als überall anwendbar vorauszusetzen, auch z.B. in der Medizin, wenn man nur "endlich" die richtigen Hierarchien und Normen finden würde - aber das ist nicht so einfach, wie sich bei genauerer Beobachtung zeigt. Es geht vielmehr darum, eine dem Sachgebiet entsprechende Art der Wissensrepräsentation zu finden.

Wenn der Mensch ein Ding künstlich konstruiert, dann kann eine Norm einfach gefunden werden, da er das Ding ja selber konstruiert hat und es deshalb Eigenschaften aufweist, die die Abstraktionen und Vorlieben des menschlichen (systematischen) Denkens widerspiegeln. Stahlschrauben und Bankkonten sind so konstruiert. Eine Norm für Briefpapier (DIN A4) lässt sich leicht definieren und in der Geschäftswelt durchsetzen. Das ist durchaus sinnvoll. Die Biologie ist aber kein Konstrukt des menschlichen Geistes. Hier lassen sich menschliche Standardforderungen weniger leicht durchsetzen. In komplexen Gebieten sind verschiedene Sichtweisen denkbar und sinnvoll. Die Medizin als im positiven Sinn empirische Wissenschaft entwickelt sich und ändert und verbessert ihre Konzepte.

*"Klassifikation ist eine Methode der Verallgemeinerung. Die Benutzung verschiedener Klassifikationen kann daher vorteilhaft sein, und jeder, ob Arzt, Pathologe oder Jurist, hat von seinem Standpunkt aus das Recht, die Krankheiten und Todesursachen so zu klassifizieren, wie es ihm zur Erleichterung seiner Untersuchungen und zur Erlangung allgemeingültiger Ergebnisse am geeignetsten erscheint."*

Der kursiv gedruckte Satz ist ein Zitat von William Farr aus dem Jahr 1856. W.Farr war der grosse Promotor und Praktiker der Gesundheitsstatistik im England des 19.Jahrhunderts und gilt als geistiger Vater der ICD. Offensichtlich haben seine Erben die Bedeutung seiner Erkenntnis nicht vergessen, denn der Satz wird zitiert im Band II (Regelwerk) der ICD-10.<sup>1</sup>

ad 4:

Das in These 4 genannte "Modell der Wissensrepräsentation" meint die Architektur der Begriffe. Es geht dabei nicht um die Aufzählung oder Definition der Begriffe, auch nicht um den Detaillierungsgrad des Modells (die Zahl der "Schubladen"), sondern um die relative Anordnung der "Schubladen", um die grundlegenden Möglichkeiten also, die Schubladen in Bezügen anzuordnen. Die einfachste Struktur ist die Hierarchie. Diese Struktur hat viele unbestreitbare Vorteile. Komplexere Strukturen können aber komplexe Sachverhalte wesentlich realitätsgerechter abbilden, ohne dass die Zahl der Schubladen ins Unendliche gesteigert werden muss. In einer Fortsetzung werde ich vier grundlegend verschiedene Architekturen des semantischen Raumes und ihre Unterschiede beschreiben:

- eindimensionale, lineare Architektur (Hierarchie, ICD-9, ICD-10)
- mehrdimensionale, unifokale Architektur (SNOMED, mit Kreuz-Stern ansatzweise, aber inkonsequent, die ICD-10)
- mehrdimensionale, multifokale, unipunktuelle Architektur (relationale Modelle, Frames, Conceptual Graphs, GRAIL des GALEN-Projektes)
- mehrdimensionale, multifokale, multipunktuelle Architektur (Begriffsmoleküle)

<sup>1</sup> Urban-Schwarzenberg, 1995, Band II, Seite 14

Wenn die komplexeren Architekturen die Realität besser modellieren können, bedeutet das nicht, dass einfachere Architekturen sinnlos sind. Hierarchien können als Darstellung einer bestimmten Sichtweise angesehen werden und sind für diesen Zweck ideal. Erst wenn verschiedene Gesichtspunkte eine Rolle spielen, entstehen Probleme. Was es nun braucht, ist eine Methode, um die vorgefundenen Primärdaten umfassend in grösstmöglicher Dataltreue und Komplexität darzustellen (z.B. mit Begriffsmolekülen) und sie anschliessend je nach Fragestellung in der entsprechenden einfacheren (z.B. hierarchischen) Struktur zu präsentieren. Dabei behält das Zwischenformat möglichst viel von der ursprünglichen Textinformation bei, doch ist es im Gegensatz zum ursprünglichen Freitext so strukturiert, dass es einer systematischen Auswertung eines klinikeigenen Data-Warehouses zugänglich ist. Vom reichen Zwischenformat aus können dann die verschiedenen behördlich verlangten Codes (ICD-10, DRG's usw.) automatisiert erstellt werden - sowie zusätzliche Auswertungen für die Klinik nach frei gewählten Gesichtspunkten erfolgen.

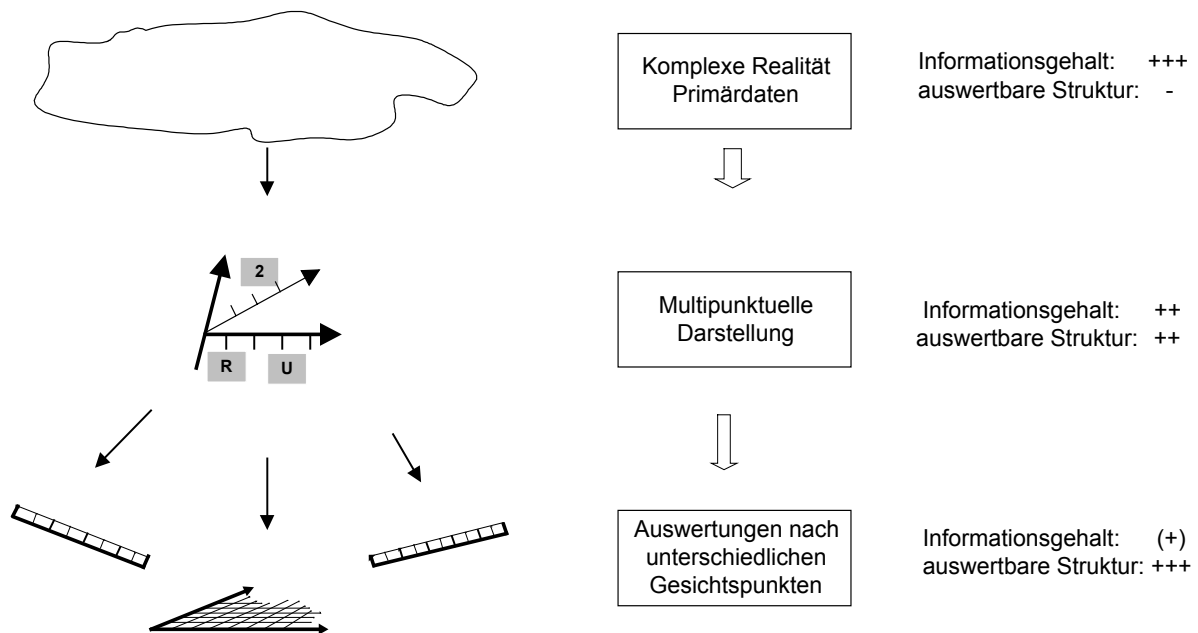


Abb 2: Zusammenspiel von komplexer und einfacher Datenrepräsentation