

2.2 Standardisierung und Komprimierung

Daten managen – 2 Datenmanipulation

Die Frage, warum Datenstandards im Gesundheitswesen so wichtig sind, beantwortet folgende Podcastfolge des NFDI Podcast mit Dr. Heinz Oberkamp mit dem Titel [„Datenstandards im Gesundheitswesen – Effizienz, Sicherheit und Transparenz“](#) aus dem Jahr 2024. Er ist über Acast, Spotify und Youtube abrufbar.

Es macht deutlich wie wichtig es ist, eine „gemeinsame Sprache“ zu sprechen, um über die eigenen Grenzen hinaus über und mit Daten kommunizieren zu können.

Kontrolliertes Vokabular

Viele Metadatenschemata nutzen kontrollierte Vokabularsysteme. Hierfür werden anstatt einer freien Verschlagwortung fest vorgegebene Terme genutzt, um die Metadaten weiter zu standardisieren. Denn selbst bei vorgegebenem Inhalt kann es zu mehrdeutigen Inhalten kommen, durch Synonyme, Singular-Plural-Problematiken oder unterschiedliche Schreibweisen. Beispielsweise könnten in einem Feld „Diagnose“ verschiedene Dinge eingetragen, aber das selbe gemeint sein, z.B. „Rückenschmerzen“ oder „schmerzender Rücken“. Es können aber auch verschiedene Detailtiefen gewählt werden wie „Rückenschmerzen“ oder „Ischiasschmerz“. Das kann schon zu Missverständnissen führen, wenn mehrere Personen gemeinsam mit einem Datensatz arbeiten, aber spätestens wenn die Daten automatisch verarbeitet werden sollen (z.B. ausgewertet werden soll, wie häufig eine bestimmte Krankheit im letzten Quartal vorgekommen ist) kommt es zu Problemen. Deswegen ist der Einsatz von kontrollierten Vokabularen sinnvoll.

Ein kontrolliertes Vokabular kann als einfache **Wortliste** realisiert sein, die den Raum der nutzbaren Begriffe beschränken, aber auch durch komplexere, strukturierte Begriffssysteme gegeben sein.

- **Taxonomie:** Eine Taxonomie erweitert eine Wortliste um Mono-Hierarchien. Das heißt, Begriffe können über Eltern-Kind-Relationen mit Ober- bzw. Unterklassen erweitert werden.
- **Thesaurus:** ein Thesaurus ist eine Taxonomie, die zusätzlich über Äquivalenz- und Assoziationsrelationen verfügt. Das bedeutet es gibt nicht nur ober- und untergeordnete Begriffe, sondern es werden auch Synonyme (manchmal auch Antonyme) aufgeführt. Das führt zu einem komplexeren Wortnetzgebilde.
- **Ontologie:** Ontologien finden vor allem in der Informatik Verwendung und enthalten zusätzlich Inferenz- und Integritätsregeln. Sie umfassen eine geordnete Darstellung einer Menge von Begriffen und der zwischen ihnen bestehenden Beziehungen, die z.B. mit logischen Relationen beschrieben werden können sowie Eigenschaften der Begriffe.

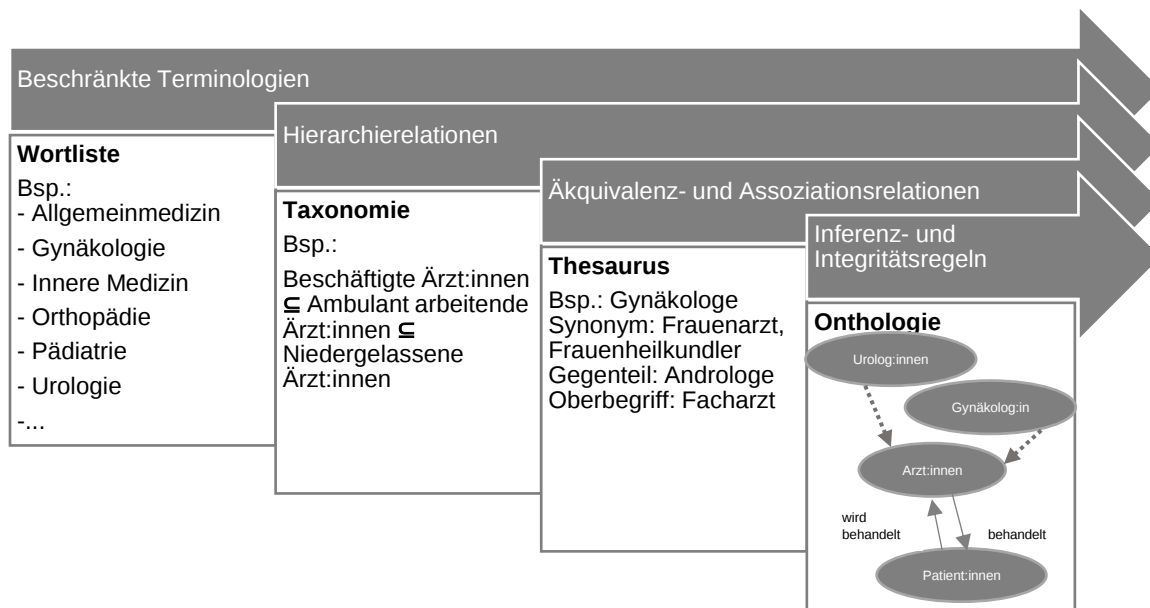


Abb. 1. Übersicht der verschiedenen Stufen kontrollierter Vokabulare

Standardisierung

Standardisierung dient dazu ein einheitliches Verständnis bei einer gemeinsamen Nutzung von Daten durch Regeln und Richtlinien zu schaffen. Die Standards legen fest wie Daten „auszusehen“ haben.

Im Gesundheitswesen existieren einige medizinische Klassifikationssysteme, wie ICD-10, OPS, aG-DRG. Diese Datenstandards sind einerseits dazu da sensible Informationen schützen, bilden aber auch die Grundlage für die Interoperabilität von Systemen und Datensätzen.

ICD-10-GM

Die ICD-Codes (International Statistical Classification of Diseases and Related Health Problems) des Diagnoseklassifikationssystems sind dazu da, Krankheitsdiagnosen zu verschlüsseln, beispielsweise auf einer Arbeitsunfähigkeitsbescheinigung. Diese Klassifikation ist global gültig und wird von der WHO (World Health Organization) herausgegeben. In Deutschland wird aktuell größtenteils die ICD-10-GM verwendet, also die 10. Entwurfsfassung in der „German Modification“, die die Besonderheiten des deutschen Gesundheitswesens berücksichtigt.

Es existiert ein systematisches Verzeichnis der ICD-10-GM, die sogenannte Systematik, welche sich in 22 Kapitel untergliedert. Unter diese fallen verschiedene Gruppen, die sich nochmal in Kategorien aufspalten und jeweils noch Subkategorien haben, die weitere Subkategorien haben können. So kommt man auf 5 Hierarchieebenen. Die ICD-10-GM Kapitel enthalten zusätzliche Klassenattribute die als nähere Hinweise, Inklusiva und Exklusiva enthalten sein können.

Im ICD-10-GM existieren auch sogenannte sekundäre Schlüsselnummern, die nur in Kombination mit einem Primärkode verwendet werden dürfen. Diese sind entweder mit einem Ausrufezeichen (um Diagnosen zu spezifizieren) oder einem Stern (um zusätzliche Manifestationen zu codieren) versehen. Zusätzlich kann man einen Code noch mit

Zusatzkennzeichen für die Diagnosesicherheit (nur für den ambulanten Bereich) und die Seitenlokalisierung versehen.

Kapitel	Gliederung	Titel
<u>I</u>	<u>A00-B99</u>	Bestimmte infektiöse und parasitäre Krankheiten
<u>II</u>	<u>C00-D48</u>	Neubildungen
<u>III</u>	<u>D50-D90</u>	Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems
<u>IV</u>	<u>E00-E90</u>	Endokrine, Ernährungs- und Stoffwechselkrankheiten
<u>V</u>	<u>F00-F99</u>	Psychische und Verhaltensstörungen
<u>VI</u>	<u>G00-G99</u>	Krankheiten des Nervensystems
<u>VII</u>	<u>H00-H59</u>	Krankheiten des Auges und der Augenanhangsgebilde
<u>VIII</u>	<u>H60-H95</u>	Krankheiten des Ohres und des Warzenfortsatzes
<u>IX</u>	<u>I00-I99</u>	Krankheiten des Kreislaufsystems
<u>X</u>	<u>J00-J99</u>	Krankheiten des Atmungssystems
<u>XI</u>	<u>K00-K93</u>	Krankheiten des Verdauungssystems
<u>XII</u>	<u>L00-L99</u>	Krankheiten der Haut und der Unterhaut
<u>XIII</u>	<u>M00-M99</u>	Krankheiten des Muskel-Skelett-Systems und des Bindegewebes
<u>XIV</u>	<u>N00-N99</u>	Krankheiten des Urogenitalsystems
<u>XV</u>	<u>O00-O99</u>	Schwangerschaft, Geburt und Wochenbett
<u>XVI</u>	<u>P00-P96</u>	Bestimmte Zustände, die ihren Ursprung in der Perinatalperiode haben
<u>XVII</u>	<u>Q00-Q99</u>	Angeborene Fehlbildungen, Deformitäten und Chromosomenanomalien
<u>XVIII</u>	<u>R00-R99</u>	Symptome und abnorme klinische und Laborbefunde, die anderenorts nicht klassifiziert sind
<u>XIX</u>	<u>S00-T98</u>	Verletzungen, Vergiftungen und bestimmte andere Folgen äußerer Ursachen
<u>XX</u>	<u>V01-Y84</u>	Äußere Ursachen von Morbidität und Mortalität
<u>XXI</u>	<u>Z00-Z99</u>	Faktoren, die den Gesundheitszustand beeinflussen und zur Inanspruchnahme des Gesundheitswesens führen
<u>XXII</u>	<u>U00-U99</u>	Schlüsselnummern für besondere Zwecke

Abb. 2. Screenshot der Kapitel des ICD-10-GM Version 2025 von <https://klassifikationen.bfarm.de/icd-10-gm/kode-suche/htmlgm2025/index.htm>

Beispiel:

- 1 Kapitel: Kapitel I Bestimmte infektiöse und parasitäre Krankheiten (A00-B99)
- 2 Gruppe: Infektiöse Darmkrankheiten (A00-A09)
- 3 Kategorie A00.- Cholera
- 4 Subkategorie: A00.0 Cholera durch *Vibrio cholerae* O:1, Biovar cholera; Klassische Cholera
- 5 Subkategorie: /

Für den Praxisalltag bietet es sich an, die Diagnosen über das alphabetische Verzeichnis, dem Alphabet, zu codieren und anhand der Systematik nochmal zu überprüfen.

OPS

Der OPS (Operationen- und Prozedurenschlüssel) wird vom BfArM jährlich herausgegeben und enthält die amtliche Klassifikation zur Verschlüsselung von Operationen, Prozeduren und allgemeine medizinischen Maßnahmen. Zusammen mit dem ICD-10-GM bildet er die Grundlage der Leistungsabrechnung im deutschen Gesundheitswesen. Auch er enthält ein systematisches und alphabetisches Verzeichnis und lässt sich in 6 Hierarchieebenen untergliedern: Kapitel, Gruppe, Kategorie und 3x Subkategorien.



Übersicht über die Kapitel

Kapitel	Gliederung	Titel
<u>1</u>	<u>1-10...1-99</u>	DIAGNOSTISCHE MASSNAHMEN
<u>3</u>	<u>3-03...3-99</u>	BILDGEBENDE DIAGNOSTIK
<u>5</u>	<u>5-01...5-99</u>	OPERATIONEN
<u>6</u>	<u>6-00...6-00</u>	MEDIKAMENTE
<u>8</u>	<u>8-01...8-99</u>	NICHT OPERATIVE THERAPEUTISCHE MASSNAHMEN
<u>9</u>	<u>9-26...9-99</u>	ERGÄNZENDE MASSNAHMEN

Abb. 3. Screenshot der Kapitel des OPS Version 2025 von [https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/ node.html](https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/node.html)

ICF

Das ICF (International Classification of Functioning, Disability and Health) ist die internationale Klassifikation der Funktionsfähigkeit, Behinderung und Gesundheit der WHO. Sie dient dazu bio-psycho-sozialen Aspekte von Krankheitsfolgen unter Berücksichtigung der Kontextfaktoren systematisch zu erfassen. Sie klassifiziert Beeinträchtigungen von Körperfunktionen, Körperstrukturen, Aktivitäten und Partizipation und Umweltfaktoren. Um die Funktionsfähigkeit, die Funktionsfähigkeit und die Behinderung eines Menschen zu beschreiben sind mehrere Codes aus allen vier Komponenten nötig und werden meist mit Beurteilungsmerkmalen ergänzt, die Informationen zu dem Ausmaß einer Schädigung, Art einer Veränderung oder Lokalisation zu geben.

ICHI

Die ICHI (International Classification of Health Interventions) ist eine Klassifikation für Gesundheitsinterventionen der WHO. Sie dazu dient, verschiedene medizinische, chirurgische, diagnostische, präventive und rehabilitative Maßnahmen zu standardisieren und international vergleichbar zu machen. Die ICHI ist in drei Achsen gegliedert: Target (Was wird behandelt?), Action (Was wird getan?) und Means (Wie wird es getan?), um eine detaillierte und standardisierte Beschreibung von Interventionen zu ermöglichen.

aG-DRG

Das aG-DRG ist ein System zur Klassifikation von Fallgruppen, dass dazu dient verschiedene Versorgungsleistungen im Krankenhaus festzulegen und zu ermitteln. Behandlungsfälle, die sich medizinisch und im Behandlungsaufwand ähneln, werden dort zu „diagnosis related groups“ zusammengefasst, als Grundlage für das leistungsorientierte Entgeltsystem im deutschen Krankenhauswesen. Ziel ist es, so mehr Leistungsgerechtigkeit und Anreize für eine effiziente Patient:innenversorgung zu erreichen. Zentrale Bausteine sind die Entgeltkataloge, die Abrechnungsbestimmungen und die Kodierrichtlinien.

Natürlich existieren noch weitere medizinische Klassifikationssysteme, beispielsweise die ATC zur anatomisch-therapeutisch-chemische Klassifikation pharmakologischer Wirkstoffe, die ICD-O-3 zur internationalen Klassifikation der Krankheiten für die Onkologie, die IVD-EDMA für die Benennung von In-vitro-Diagnostika

Konvertierung

Eine Konvertierung ist die Überführung von einem Datenformat in ein anderes. Beispielsweise kann eine Konvertierung sinnvoll sein, wenn eine Datei in einem nicht-offenen Format vorliegt und in ein offenes überführt werden soll. Aber Achtung! Bei der Konvertierung von Dateien können Informationen verloren gehen. Der Informationsverlust kann sich entweder durch einen Verlust der Qualität äußern, aber auch Metadaten können verloren gehen. Bei statistischen Daten können Wertedefinitionen oder Variablenlabels



verschwinden und bei textuellen Daten gehen häufig Bearbeitungen wie Hervorhebungen oder Kopf- und Fußzeilen verloren. Deswegen sollten mögliche Konvertierungsschritte immer dokumentiert werden und wenn möglich auch die Originaldateien behalten werden.

Komprimierung

Komprimierung ist ein Prozess, der dafür gedacht ist, die Dateigröße von Daten und Dateien zu verringern. Das ist sinnvoll, um Speicherplatz zu sparen und Übertragungsgeschwindigkeiten zu erhöhen.

Datentypen	Nicht-komprimierte Formate
Text	TXT, XML, PDF/A, HTML, ASCII, UTF-8
Bild	TIFF, JPEG 2000, PDF, PNG
Video	MOV, MPEG, AVI, MXF
Sounds	WAVE, AIFF
Statistik	ASCII
Datenbanken	XML, CSV

Man unterscheidet zwischen verlustfreien und verlustbehafteten Komprimierung. Bei der verlustfreien Komprimierung (auch Redundanzreduktion genannt) werden in erster Linie Redundanzen, also überflüssige Bestandteile, entfernt. Wichtig ist, dass aus den verlustfrei komprimierten Daten durch ein Dekomprimierungsverfahren exakt die Ursprungsdatei wiederhergestellt werden könnte. Beispielsweise ist ein häufig in der Praxis genutztes Komprimierungsformat die ZIP-Datei. Diese nutzt Mechanismen wie die Wörterbuchmethode und Entropiekodierung um Redundanzen (vor allem aus Texten) zu entfernen.

Bei der verlustbehafteten Komprimierung, oder Irrelevanzreduktion, geht immer ein Teil der Informationen verloren. Die Algorithmen versuchen diesen Verlust aber so wenig wahrnehmbar wie möglich zu machen. Verlustbehaftete Komprimierung passiert vor allem bei Bildern, Video- und Audiodateien. Hier macht man sich die Grenzen der menschlichen Wahrnehmung zu Nutze und entfernt beispielsweise Frequenzen, die für Menschen nicht hörbar sind und reduziert Farbaufösungen, die das menschliche Auge nicht differenzieren kann.



Quellenverzeichnis

Bundesinstitut für Arzneimittel und Medizinprodukte (2025). Aufbau der Vierstelligen ausführlichen Systematik der ICD-10-GM. Abgerufen am 30.09.2025, von <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/Systematik/systematikaufbau.html>

Bundesinstitut für Arzneimittel und Medizinprodukte (2025). ICD-10-GM. Abgerufen am 30.09.2025, von <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/ node.html>

Bundesinstitut für Arzneimittel und Medizinprodukte (2025). ICF. Abgerufen am 30.09.2025, von <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICF/ node.html>

Bundesinstitut für Arzneimittel und Medizinprodukte (2025). ICHI. Abgerufen am 30.09.2025, von <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/ICHI/ node.html>

Bundesinstitut für Arzneimittel und Medizinprodukte (2025). Klassifikationen. Abgerufen am 30.09.2025, von <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ node.html>

Lamar Soutter Library & University of Massachusetts Medical School (2022). New England Collaborative Data Management Curriculum: - 02 Types, Formats, and Stages of Data. Abgerufen am 30.09.2025, von <https://oercommons.org/courses/new-england-collaborative-data-management-curriculum>

MIT Libraries Data Management Group (2016). Research Data Management: Strategies for Data Sharing and Storage. Abgerufen am 30.09.2025, von <https://www.twillo.de/edu-sharing/components/render/01887843-1814-42e6-9ad5-6941a88d5c5a>

Spitzenverband Bund der Krankenkassen (2025). aG-DRG-System. Abgerufen am 30.09.2025, von https://www.gkv-spitzenverband.de/krankenversicherung/krankenhaeuser/drg_system/ag_drg_system_2025/ag_drg_2025.jsp

Sablatnig, R., Zambanini, S., & Licandro, R. (2014). Einführung in Visual Computing. Kapitel 05 Bildcodierung und Kompression. Abgerufen am 30.09.2025, von <https://www.cg.tuwien.ac.at/courses/EinfVisComp/Skriptum/SS14/EVC-05%20Bildcodierung%20und%20Kompression.pdf>

WHO & BfArM (2021). ICD-10-GM Version 2022: Anleitung zur Verschlüsselung. Abgerufen am 30.09.2025, von <https://klassifikationen.bfarm.de/icd-10-gm/kode-suche/htmlqm2022/zusatz-04-anleitung-zur-verschluesselung.htm>

World Health Organization (2025). International Classification of Health Interventions (ICHI). Abgerufen am 30.09.2025, von <https://www.who.int/classifications/international-classification-of-health-interventions>

Weiterführende Info

Bundesinstitut für Arzneimittel und Medizinprodukte (2025). Aufbau der Vierstelligen ausführlichen Systematik der ICD-10-GM. Abgerufen am 30.09.2025, von <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/Systematik/systematikaufbau.html>

DIMDI (2010). Basiswissen Kodieren, Eine kurze Einführung in die Anwendung von ICD-10-GM und OPS. Abgerufen am 30.09.2025,



von https://www.bfarm.de/SharedDocs/Downloads/DE/Kodiersysteme/basiswissen-kodieren-2010.pdf?__blob=publicationFile

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt



**Finanziert von der
Europäischen Union**
NextGenerationEU

Das Material ist im Rahmen des Projekts „DIM.RUHR: Datenkompetenzzentrum für die interprofessionelle Nutzung von Gesundheitsdaten in der Metropole Ruhr“ (Förderkennzeichen: 16DKZ2008[A-F]) entstanden. Dementsprechend spiegelt das Material in Version 1.0 den Stand von 2025 wider.

„Daten managen - 2.2 Datenmanipulation – Standardisierung und Komprimierung“
von Anne Mainz und Sven Meister.

Dieses Werk und dessen Inhalt sind – sofern nicht anders angegeben – lizenziert unter CC BY 4.0. Ausgenommen aus der Lizenz sind die verwendeten Logos.

