



hochschule mannheim

in Kooperation mit



miracum

Biomedizinische Informatik und Data Science (M.Sc.)

Master of Science – Zertifikatskurse/-programme

PROJEKTARBEIT

Technische Abbildung des Lebenszyklus von ICD-10-GM Klassifikationen

Abel HODELÍN HERNÁNDEZ 

betreut von
Marcus WILL

15. November 2021

Inhaltsverzeichnis

Abbildungsverzeichnis	I
Tabellenverzeichnis	II
1 Überblick	6
1.1 Einleitung	6
1.2 Verzeichnisse der ICD-10-GM Klassifizierung	6
1.2.1 Alphabetisches Verzeichnis (Alphabet)	7
1.2.2 Systematisches Verzeichnis (Systematik)	7
1.3 Metadaten	9
2 Materialien und Methode	10
2.1 Aufbau der DB	10
2.2 Funktion der Tabellen	10
2.2.1 Tabelle „kodes“ von BfArM	10
2.2.2 Neue Tabellen	11
2.3 Fluss der Information	12
2.4 ETL-Strecke	12
2.4.1 Extraktion	14
2.4.2 Transformation	15
2.4.3 Laden	15
3 Ergebnisse	16
3.1 ICD-10-GM in der CSV-Dateien und DB	16
3.2 Zeitliche Entwicklung der ICD-10-GM	17
3.3 Neue ICD-10-GM	18
3.4 Modifizierte ICD-10-GM	21
3.5 Gelöschte ICD-10-GM	24
3.6 Wiederverwendete ICD-10-GM	26
3.7 Strukturelle Änderungen	26
4 Schlussfolgerung	28
4.1 Weitere Nutzungen des Systems	28

Abbildungsverzeichnis

2.1	Datenbankstruktur	11
2.2	Datenfluss des Prozesses	13
2.3	ETL-Strecke	14
3.1	ICD-10-GM in der DB	17
3.2	Neue ICD-10-GM pro Jahr	19
3.3	Kapitel mit den meisten eingeführten ICD-10-GM (2008 - 2021) . . .	20
3.4	Neue meldepflichtige ICD-10-GM pro Jahr	20
3.5	Gelöschte ICD-10-GM pro Jahr	24
3.6	Meist betroffene Kapitel beim Löschen 2013	25
3.7	Hierarchie der gelöschten ICD-10-GM	25

Tabellenverzeichnis

1.1	Verschiedene Bezeichnungen von COVID-19	7
1.2	Beispiele von Kapiteln.	8
1.3	Beispiele von Gruppen.	8
1.4	Beispiele von Kategorien und Subkategorien.	9
1.5	Liste der Dateien im Ordner Klassifikationsdateien.	9
3.1	ICD-10-GM in den CSV-Dateien	16
3.2	Beispiel der Historisierung einer ICD-10-GM Klassifikation	18
3.3	Meldepflichtige ICD-10-GM	21
3.4	Änderungen in den Metadaten	22
3.5	Beispiel von seltenen Krankheiten in Mitteleuropa (Diphtherie und Poliomyelitis)	23
3.6	Wieder benutzte ICD-10-GM	26

Abkürzungsverzeichnis

Alpha-ID	Alpha-Identifikator
BASH	Bourne Again Shell
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
BMG	Bundesministerium für Gesundheit
cDW	Clinical Data Warehouse
COVID	Corona Virus Disease
CSV	Comma-Separated Values
DB	Datenbank
DIZ	Datenintegrationszentrum
EBM	einheitlicher Bewertungsmaßstab
ETL	Extract, Transform, Load
G-DRG	German Diagnosis Related Groups
ICD-10-GM	International Statistical Classification of Diseases and Related Health Problems, 10. Revision, German Modification
ISO	International Organization for Standardization
KRINKO	Kommission für Krankenhaushygiene und Infektionsprävention
MII	Medizin Informatik Initiative
MIRACUM	Medical Informatics in Research and Care in University Medicine
RDBMS	Relational Database Management System
RKI	Robert-Koch-Institut

SGB	Sozialgesetzbuch
SQL	Structured Query Language
UTF	Universal Coded Character Set Transformation Format
WHO	Weltgesundheitsorganisation

1 Überblick

1.1 Einleitung

Die International Statistical Classification of Diseases and Related Health Problems, 10. Revision, German Modification (ICD-10-GM) auf Deutsch Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme gilt als Standard bei der Aufzeichnung von Gesundheitsdaten. Häufig bleibt die temporäre Natur dieses Schemas in Applikationen und Auswertungen unberücksichtigt.

Im Rahmen der Konzeption des Clinical Data Warehouse (cDW) für ein Datenintegrationszentrum (DIZ) des Konsortiums Medical Informatics in Research and Care in University Medicine (MIRACUM) der Medizin Informatik Initiative (MII) muss diese Situation abgebildet werden.

Das Ziel dieses Projekts ist die Darstellung der Besonderheiten eines Lebenszyklus der historischen ICD-10-GM Fassungen von 2007 bis 2021. Um dieses Ziel zu erreichen wurde ein Datenbank (DB) Schema für die Speicherung der Information des semantischen Verzeichnisses der ICD-10-GM entwickelt, eine Extract, Transform, Load (ETL)-Strecke konzipiert, um die Daten in die DB zu importieren; und Structured Query Language (SQL)- und Python-Skripts, um die Daten zu analysieren.

1.2 Verzeichnisse der ICD-10-GM Klassifizierung

Die ICD-10-GM ist die amtliche Klassifikation zur Verschlüsselung von Diagnosen in der ambulanten (§295 SGB V) und stationären (§301 SGB V) Versorgung in Deutschland, deren Versionen und Formate gelten vom Anfang bis zum Ende eines Jahres und werden im Auftrag des Bundesministeriums für Gesundheit (BMG) von dem Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) jährlich aktualisiert und herausgegeben [1].

Diese Klassifikation ist alphanumerisch, monohierarchisch strukturiert mit bis zu 5 Ebenen und deren Aufbau besteht aus zwei Teilen, das systematische und alphabetische Verzeichnis [1].

1.2.1 Alphabetisches Verzeichnis (Alphabet)

Die alphabetische Zuordnung des Codes entsteht aus den gewohnten Texten für die Diagnosen. Die Systematik enthält nicht alle Diagnosen des Alphabets. Deswegen dient die semantische Bezeichnung als Verzugsbezeichnung, sodass das Alphabet auch andere veraltete oder ungenaue Diagnosenbezeichnungen enthält [2]. Die Tabelle 1.1 illustriert ein Beispiel dieser Ungenauigkeit anhand von der Diagnose Corona Virus Disease (COVID-19). Aus diesem Grund existiert der Alpha-Identifikator (Alpha-ID). Der nutzt die Bezeichnungen als Basis für einen nicht klassifizierenden Kode [2].

Tabelle 1.1: Verschiedene Bezeichnungen von COVID-19 mit deren Alpha-ID und ICD-10-GM Identifikatoren

Alpha-ID	ICD-10-GM	Bezeichnung
I130805	U07.1!	Coronavirus-Infektion-2019, durch Labortest nachgewiesen
I130804	U07.1!	Coronavirus-Infektion-2019, Virus nachgewiesen
I130797	U07.1!	Coronavirus-Krankheit-2019, Virus nachgewiesen
I130809	U07.1!	COVID-19-Infektion, durch Labortest nachgewiesen
I130796	U07.1!	COVID-19-Infektion, Virus nachgewiesen

1.2.2 Systematisches Verzeichnis (Systematik)

Die Systematik ist die hierarchisch geordnete Liste der ICD-10-GM Kodierung [3]. Diese Hierarchie enthält bis zu fünf Ebenen, diese sind in Kapitel, Gruppen, Kategorien (Kode bis zur dritten Stelle), und Subkategorien (vierstellige oder fünfstellige Codes) gegliedert [4].

Für die Durchführung und Zwecke dieses Projekts wird mit dem systematischen Verzeichnis gearbeitet.

Kapitel

Der thematische Bereich aller Diagnosen ist in 22 Kapiteln, von I bis XXII, abgedeckt. Der Kern davon sind die organspezifischen Krankheiten von Kapiteln III bis XIV. Die nicht organspezifischen Krankheiten werden in den restlichen Kapiteln zusammengefasst. Jeder dargestellte Codebereich wird durch ein bis vier Buchstaben gekennzeichnet, diese Buchstaben sind das erste Zeichen einer Kodierung [4]. Die Tabelle 1.2 zeigt drei Beispiele von Kapiteln.

Tabelle 1.2: Beispiele von Kapiteln.

Kapitel	Kodebereich	Titel
I	A09-B99	Infektiöse Darmkrankheiten
II	C00-D48	Neubildungen
XX	V01-Y98	Äußere Ursachen von Morbidität und Mortalität

Gruppen

Jede Gruppe besteht aus einem unteren und oberen Kodebereich, somit werden ca. 241 Gruppen aufgebaut. Die Bereiche werden durch einen dreistelligen alphanumerischen Kode repräsentiert. Der Buchstabe der unteren und oberen Kodebereiche ist immer derselbe [4]. Die Tabelle 1.3 stellt Beispiele von drei Gruppen dar.

Tabelle 1.3: Beispiele von Gruppen.

Gruppe von	Gruppe bis	Kapitel	Titel
A00	A09	I	Infektiöse Darmkrankheiten
F99	F99	V	Nicht näher bezeichnete psychische Störungen
U98	U99	XXII	Belegte und nicht belegte Schlüsselnummern

Kategorien und Subkategorien

Kategorien oder ICD-10-GM Kodierungen haben in der Regel Subkategorien und sind maximal fünfstellig. Jeder Kodierung beginnt mit einem Großbuchstaben gefolgt von zwei Ziffern, die sogenannten dreistelligen Codes. Andere haben einen Punkt direkt nach den zwei Ziffern, gefolgt von einer oder zwei weiteren Ziffern, die vier- und fünfstelligen Codes. Manche dreistellige Codes stehen für einzelne Diagnosen und stellen somit eine Kategorie dar. Andere besitzen Subkategorien, deren Codes vier oder fünf Stellen haben können, und mehrere Krankheiten mit gemeinsamen Merkmalen zusammenfassen [4, 5].

Einige vier- oder fünfstellige Subkategorien sind postkombinierte Codes, das heißt, der Kode wird mit Hilfe einer Liste mit möglichen Werten für die restlichen Stellen gebildet [5]. Es gibt weitere Merkmale der Kategorie und Kodierung, aber diese werden nicht in diesem Projekt behandelt. Die Tabelle 1.4 illustriert Beispiele von Kategorien und Subkategorien.

Tabelle 1.4: Beispiele von Kategorien und Subkategorien.

Ebene	Stellen	Kode	Titel
Kategorie	drei	B25	Zytomegalie
Subkategorie	vier	B25.0	Pneumonie durch Zytomegalieviren
Subkategorie	fünf	B25.80	Infektion des Verdauungstraktes durch Zytomegalieviren

1.3 Metadaten

Das BfArM veröffentlicht zwischen September und Dezember eines Jahres die Metadaten der neuen Fassung der ICD-10-GM in einem komprimierten Ordner in einer ZIP-Datei mit Unterordnern mit Kodierungsdateien und weiteren Informationen. Der Unterordner Klassifikationsdateien enthält die Dateien der aktuellen Fassung der ICD-10-GM wie in der Tabelle 1.5 dargestellt wird. Diese Dateien nutzen das Comma-Separated Values (CSV)-Format mit Semikolon als Trennzeichen. Die ZIP-Datei enthält auch eine Liesmich-Datei unter dem Namen „icd10gmJJJJsys_tmetadaten_liesmich.txt“, wobei „JJJJ“ das Jahr der Fassung darstellt, mit der Beschreibung aller CSV-Dateien, Information der vorkommenden Erneuerungen in der Struktur dieser Dateien, und SQL-Statements für den Aufbau einer DB. Die ZIP-Dateien der Version des laufenden Jahres und der Vorgängerversionen befinden sich auf der Seite Downloads der Klassifikationen der BfArM Webseite in der Sektion ICD-10-GM. Die Namen der Links zu den ZIP-Dateien haben die Struktur „ICD-10-GM Version JJJJ“ und die Struktur der Namen der ZIP-Dateien ist „ICD-10-GMJJJJ.zip“.

Tabelle 1.5: Liste der Dateien im Ordner Klassifikationsdateien.

Dateiname	Information
icd10gmJJJJsys_gruppen.txt	Gruppen der ICD-10-GM-Systematik
icd10gmJJJJsys_kapitel.txt	Kapitel der ICD-10-GM-Systematik
icd10gmJJJJsys_kodes.txt	Kodes der ICD-10-GM-Systematik
morbl_JJJJ.txt	Morbiditätsliste
mortl1_JJJJ.txt	Mortalitätsliste 1
mortl1grp_JJJJ.txt	Gruppen der Mortalitätsliste 1
mortl2_JJJJ.txt	Mortalitätsliste 2
mortl3_JJJJ.txt	Mortalitätsliste 3
ortl3grp_JJJJ.txt	Gruppen der Mortalitätsliste 3
mortl4_JJJJ.txt	Mortalitätsliste 4

2 Materialien und Methode

2.1 Aufbau der DB

Für die Durchführung dieses Projekts wurde ein relationales DB Schema in PostgreSQL entwickelt. Die Wahl von PostgreSQL für dieses Projekt beruht auf dem verbreiteten Einsatz dieses Relational Database Management System (RDBMS) im MIRACUM Konsortium. Zusätzlich ist PostgreSQL ein freies Open Source RDBMS mit vielen Features [6].

Für die Implementierung der DB wurden die SQL-Anweisungen, nach Empfehlung in der Liesmich-Datei der Fassungen, angepasst [7]. Außerhalb von den vorher genannten Anpassungen wurde die Tabelle „kodes“ auch erweitert und modifiziert. In dieser Tabelle wurde die Spalte „ver“ für die Speicherung der Fassungen oder Versionen eingefügt und die Spalte „fünfsteller“ wurde zu „fuenfsteller“ umbenannt, um Probleme mit der Zeichen-Kodierung zu verhindern. Diese Veränderungen sind schwarz in der Tabelle „kodes“ der Abbildung 2.1 gekennzeichnet. Noch dazu wurden drei neue Tabellen eingefügt, um die Veröffentlichung Daten („icd10gm_release_info“), Speicherung („icd10gm“) und Historisierung („icd10gm_history“) aller verfügbaren Schlüsselnummern von 2007 bis 2021 zu steuern. Solche Tabellen sind mit gelben Kopfzeilen in der Abbildung 2.1 repräsentiert.

2.2 Funktion der Tabellen

2.2.1 Tabelle „kodes“ von BfArM

Die Tabelle „kodes“ vom BfArM-Schema beinhaltet die semantische Information der CSV-Dateien von nur einer Fassung und enthält damit die Anfangsdaten für die Tabellen „icd10gm“ und „icd10gm_history“.

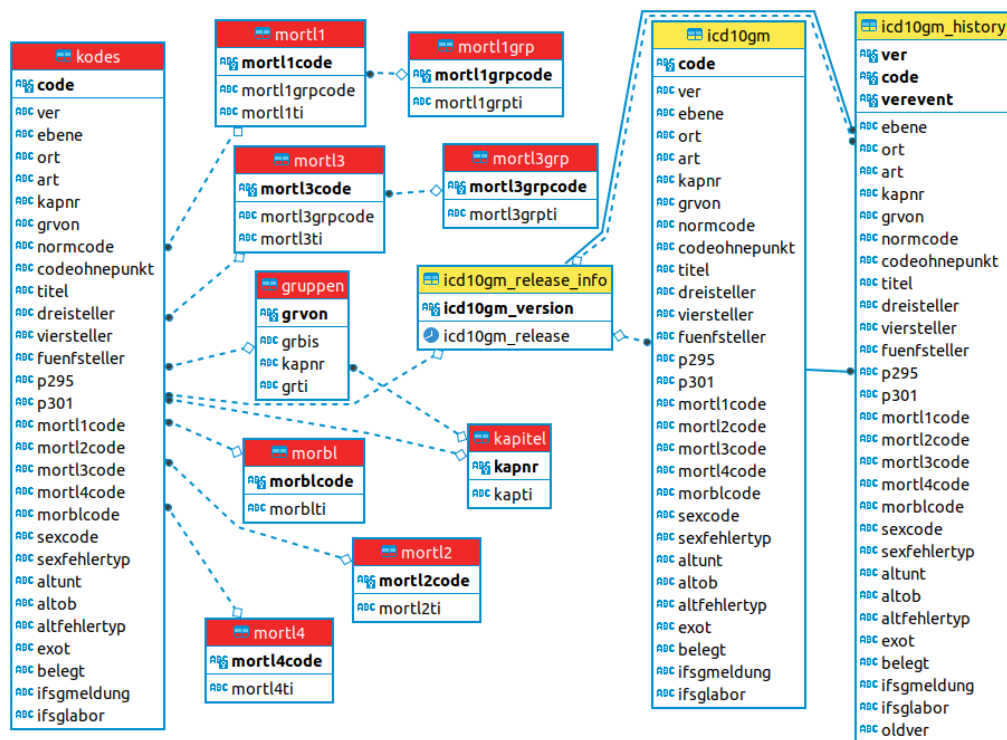


Abbildung 2.1: Datenbankstruktur für die Steuerung der ICD-10-GM. Die Tabellen von BfArM sind mit Kopfzeilen in rot gekennzeichnet, die Spalten in schwarz in diesen Tabellen sind die vorgenommenen Erweiterungen und Modifikationen. Die Tabellen mit Kopfzeilen in gelb sind die neuen eingeführten Tabellen für die Historisierung der ICD-10-GM.

2.2.2 Neue Tabellen

Für die Erfassung der Datierung der Veröffentlichungen der Versionen wurde die Tabelle „icd10gm_release_info“ erstellt. Diese speichert den Identifikator oder das Jahr der Version in der Spalte „icd10gm_version“ als Hauptschlüssel und das Datum der Veröffentlichung in der Spalte „icd10gm_release“. Diese hat das Format „JJJJ-MM-TT“. Wobei „J“ das Jahr, „M“ den Monat und „T“ den Tag darstellen.

Die ICD-10-GM von 2007 bis 2021, werden in der Tabelle „icd10gm“ gespeichert. In dieser Tabelle wird die Information der noch gültigen ICD-10-GM aktualisiert. Die gelöschten ICD-10-GM beinhalten die Information der letzten Aktualisierung. Die Spalte „code“ dieser Tabelle ist der Hauptschlüssel, genau wie in der Tabelle „kodes“ vom BfArM-Schema. Das Feld „ver“ speichert den Identifikator der Version

der Insertion oder der letzten Änderung und ist ein Fremdschlüssel, der zu der Spalte „`icd10gm_version`“ der Tabelle „`icd10gm_release_info`“ zeigt.

Die Tabelle „`icd10gm_history`“ enthält die Information der ICD-10-GM, die im Laufe der Zeit eingefügt, gelöscht oder geändert wurden. Die Besonderheiten dieser Tabelle sind die Spalten „`ver`“, „`oldver`“ und „`verevent`“. Die Identifikatoren vergangener Versionen einer ICD-10-GM werden in der Spalte „`oldver`“ gespeichert. Die Spalte „`ver`“ enthält den Identifikator der Fassungen bei deren eine ICD-10-GM eingefügt, gelöscht oder modifiziert wurde. Die Ereignisse einer Insertion, Änderung, Löschung oder Wiederverwendung werden mit den Buchstaben „I“ (insert), „U“ (update), „D“ (delete) und „DI“ (delete insert) in der Spalte „`verevent`“ kodiert und gespeichert.

2.3 Fluss der Information

Mit Hilfe einer ETL-Strecke werden die Daten aus der ZIP-Datei in der DB importiert. Die Information der Veröffentlichung der ICD-10-GM wird in der Tabelle „`icd10gm_release_info`“ eingefügt. Bei jedem Laden wird die Information der Tabelle „`kodes`“ vorher gelöscht und mit neuen Datensätzen geladen.

Die Codes, die vorher nicht in der Tabelle „`icd10gm`“ vorhanden waren, werden in diese und in die Tabelle „`icd10gm_history`“ kopiert. Diese Datensätze werden in „`icd10gm_history`“ als eingefügt markiert.

Nicht in „`kodes`“ existierende ICD-10-GM werden aus „`icd10gm`“ in „`icd10gm_history`“ kopiert. Diese Kopie wird als gelöscht markiert, und enthält die Indikatoren der Fassung der Insertion oder der letzten Änderung in der Spalte „`oldver`“, und der Fassung der Löschung in der Spalte „`ver`“.

Existierende Kodierungen mit modifizierten Informationen werden in der Tabelle „`icd10gm`“ aktualisiert und nachher in „`icd10gm_history`“ hinzugefügt. In diesem Fall werden auch die Indikatoren der alten und neuen Fassung registriert und der Datensatz als modifiziert markiert.

Dieser Fluss der Information ist im Datenflussdiagramm der Abbildung 2.2 repräsentiert. Der automatisierte Prozessablauf in der DB ist durch Trigger in den Tabellen „`kodes`“ und „`icd10gm`“ gesteuert.

2.4 ETL-Strecke

Für den Import der Information der ICD-10-GM in der DB wurde eine ETL-Strecke entwickelt. Dazu wurden Bourne Again Shell (BASH)-Skripts unter Ubuntu 20.04 programmiert und durchgeführt. Die Entscheidung von BASH basiert sich darauf, dass BASH die Standard interaktive Shell und Skript Sprache von Linux-Betriebssystemen

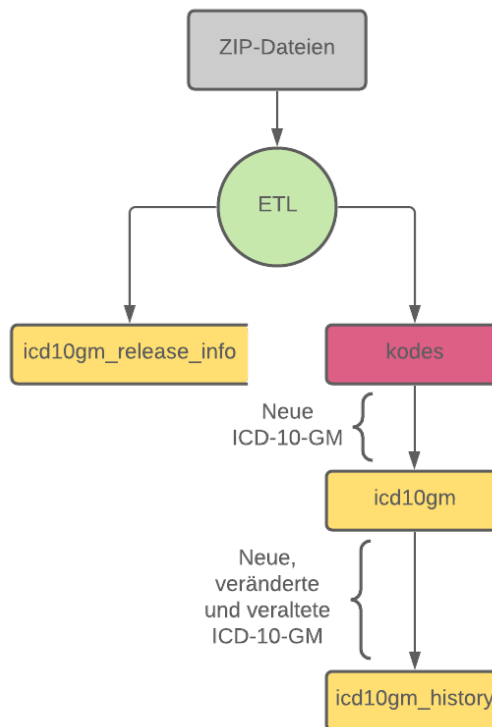


Abbildung 2.2: Datenflussdiagramm der Informationen von den ZIP-Dateien bis zum Import in der DB.

ist und ermöglicht auch die Automatisierung einer Reihenfolge von Kommandos des Betriebssystems [8]. Da die Interaktion mit den Befehlen des Betriebssystems via BASH effizienter und leichter ist, ist es möglich in einer sicheren Art die Zugangsdaten für die DB zu nutzen, ohne diese Information in Klartext in den Skripten oder in anderen Dateien sichtbar zu platzieren.

Der Anfangspunkt für die Durchführung der ETL sind die ZIP-Dateien der ICD-10-GM Metadaten von 2007 bis 2021 aus der Download Seite vom BfArM für die Klassifikationen. Diese Dateien wurden zuerst manuell heruntergeladen. Der Abruf und die Reihenfolge der Skripts für den Durchlauf jedes Schritts der ETL werden von einem zentralen BASH-Skript definiert. Die Abbildung 2.3 stellt das Flussdiagramm der ETL-Strecke dar.

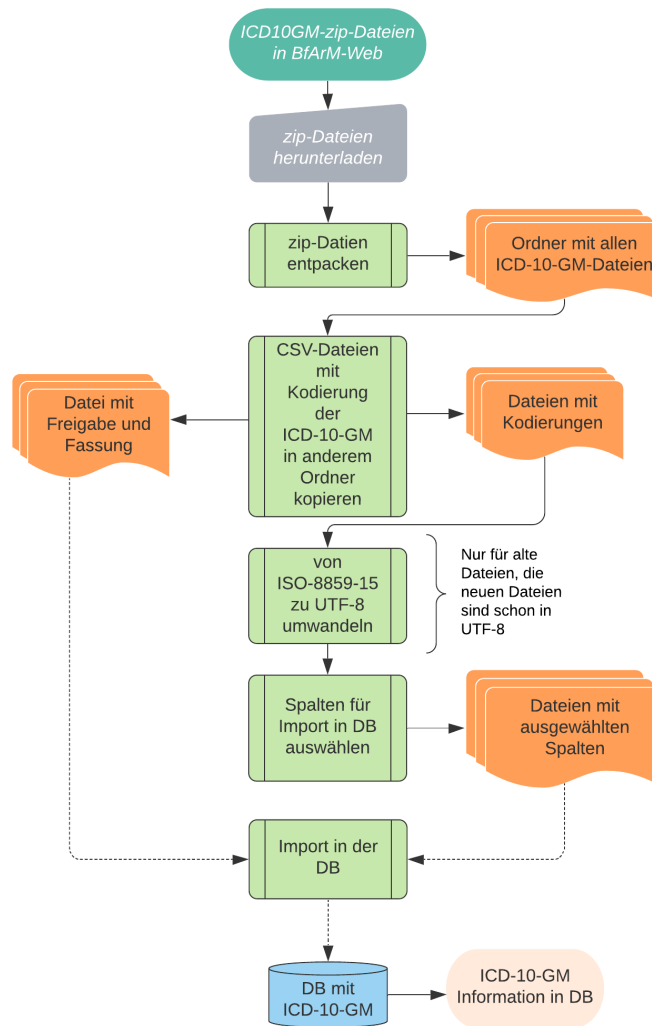


Abbildung 2.3: Flussdiagramm der ETL-Strecke für den Import der Informationen der ICD-10-GM aus den ZIP-Dateien in der DB.

2.4.1 Extraktion

Zuerst werden die Ordner der ZIP-Dateien entpackt und ein neuer Ordner für die Speicherung der Code-Dateien wird generiert. Diese Dateien werden ausgewählt und in den hergestellten Ordner kopiert. Die Information des Datums der Freigabe und Fassung wird aus der Liesmich-Datei extrahiert und in eine CSV-Datei importiert.

2.4.2 Transformation

Die Kode-Dateien von 2007 bis 2009 haben den Windows-Standardzeichensatz International Organization for Standardization, ISO-8859-15 als Zeichenkodierung. Das verursacht Probleme bei dem Datenaustausch zwischen Plattformen, da die PostgreSQL Instanz in Universal Coded Character Set Transformation Format, UTF-8 konfiguriert ist. Aus diesem Grund wird das Format dieser Kode-Dateien von ISO-8859-15 in UTF-8 umgewandelt.

Im Laufe der Zeit sind neue Spalten in der CSV-Dateien der Klassifizierungen entstanden und andere Felder werden veraltet und entnommen [9, 10]. Deswegen werden die aktuell benutzten Spalten ausgewählt, leere Felder werden an den Positionen der Spalten, die vorher nicht vorhanden waren, eingefügt. Ein neues Feld mit der Version am Anfang jeder Zeile ist eingeführt, und der Datensatz mit den Änderungen in einer neuen generierten CSV-Datei gespeichert.

2.4.3 Laden

Am Ende der Transformation wird die Information der zuletzt generierten CSV-Dateien mit Kodes in die Tabelle „kodes“ der DB geschrieben. Der interne Datenfluss in der DB ist in der Sektion 2.3 beschrieben.

3 Ergebnisse

3.1 ICD-10-GM in der CSV-Dateien und DB

Die Analyse der Information der ICD-10-GM wurde mit Hilfe von SQL-Befehlen und der Skript Sprache Python unter dem Framework Jupyter-Notebook durchgeführt.

Wie in der Tabelle 3.1 dargestellt wird, enthält jede Kode-Datei mehr als 15450 Codes. Obwohl manche Kodierungen gelöscht werden, nimmt die Menge neuer ICD-10-GM ständig zu.

Tabelle 3.1: Anzahl an ICD-10-GM pro Jahr in den CSV-Dateien

Anzahl	Fassung
15455	2007
15498	2008
15523	2009
15598	2010
15633	2011
15643	2012
15668	2013
15688	2014
15761	2015
15821	2016
15930	2017
16059	2018
16126	2019
16131	2020
16203	2021
236737	Gesamt

Nach dem Durchlauf der ETL sind insgesamt 16520 ICD-10-GM von 2007 bis 2021 in der DB eingefügt. In der Abbildung 3.1 ist das Verhältnis dieser Kodierungen in Bezug auf die Modifikationen dargestellt. Die konkreten Zahlen dieser Modifikationen sind in der Tabelle 3.4 dargestellt.



Abbildung 3.1: Charakterisierung der ICD-10-GM in der DB in Bezug auf die Modifikationen.

3.2 Zeitliche Entwicklung der ICD-10-GM

Die Darstellung der Besonderheiten eines Lebenszyklus der historischen ICD-10-GM Fassungen ist in der Tabelle 3.2 exemplarisch dargestellt. Das Beispiel zeigt die chronologische Entwicklung der ICD-10-GM in der Tabelle „icd10gm_history“ der DB. Die Spalte Fassung, „ver“ in der Tabelle „icd10gm_history“, zeigt die Version der Einführung, Änderung, Löschung oder Wiederverwendung einer ICD-10-GM. Das Feld Alte Fassung, „oldver“ in der vorher genannten Tabelle, beinhaltet den Identifikator der vorherigen Fassung einer Schlüsselnummer. In der Spalte Ereignis, „ver-event“ in der Tabelle „icd10gm_history“, sind die Ereignisse Einführung, Änderung, Löschung und Wiederverwendung, wie in der Subsektion 2.2.2 beschrieben wurden, kodiert. Eine wichtige Anmerkung in der Darstellung ist, dass die gezeigte alte Fassung bei der wiederverwendeten Klassifikationen, die Version der Insertion oder der letzten Änderung ist, weil dieser Code in der Fassung der Löschung nicht vorhanden war.

Somit könnten wir erkennen, dass unsere Implementierung die Lebenszyklus von ICD-10-GM Klassifikationen darstellen kann, sodass jede Modifikation in der Semantik zwischen 2007 und 2021 dokumentiert ist. In den weiteren Sektionen werden die verschiedenen Änderungen in dem semantischen Verzeichnis und einigen Sonderverzeichnissen gezeigt und deren Gründe erklärt.

Tabelle 3.2: Beispiel der Historisierung einer ICD-10-GM Klassifikation. Diese stellt eine, nach der Einführung, veränderte, gelöschte, und wieder benutzte Schlüsselnummer dar.

Fassung	Kode	Titel	Alte Fassung	Ereignis
2007	N90.8	Sonstige näher bezeichnete nicht-entzündliche Krankheiten der Vulva und des Perineums		I
2008	N90.8	Sonstige näher bezeichnete nicht-entzündliche Krankheiten der Vulva und des Perineums	2007	U
2013	N90.8	Sonstige näher bezeichnete nicht-entzündliche Krankheiten der Vulva und des Perineums	2008	U
2014	N90.8	Sonstige näher bezeichnete nicht-entzündliche Krankheiten der Vulva und des Perineums	2013	D
2016	N90.8	Sonstige näher bezeichnete nicht-entzündliche Krankheiten der Vulva und des Perineums	2013	DI

3.3 Neue ICD-10-GM

Von 2008 bis 2021 sind insgesamt 966 neue ICD-10-GM entstanden. In der Abbildung 3.2 ist ein interessanter Aspekt davon repräsentiert, nämlich die Anzahl neuer ICD-10-GM pro Jahr ist unregelmäßig. Es gib Jahre wie 2017 und 2018 mit mehr als 120 neuen Einträgen. Dieses Phänomen passiert am meisten beim Bedarf neuer Subklassifikationen, um die Diagnosen spezifischer zu kodieren und an diverse klinische Systeme anzupassen. Ein Beispiel dieses Phänomens ist der Code **U81! Bakterien mit Multiresistenz gegen Antibiotika**, dieser wurde 2017 entnommen und stattdessen entstand die Kodierung **U81.-! Gramnegative Erreger mit bestimmten Antibiotikaresistenzen, die besondere therapeutische oder hygienische Maßnahmen erfordern** zusammen mit 43 weiteren Subklassifikationen von **U81.0-! bis U81.8!**. Die Ursache davon war eine Umstrukturierung der Bereiche **U81!**, um diese Codes an die Nomenklatur der Kommission für Krankenhaushygiene und Infektionsprävention (KRINKO) anzupassen [11].

Im Jahr 2018 wurden mehr als 80 neue ICD-10-GM in den Bereichen **M14.-* Arthropathien bei sonstigen anderenorts klassifizierten Krankheiten** eingefügt. Diese Zunahme wird auch in der Abbildung 3.3 widerspiegelt, denn diese Diagnosen sind im Kapitel Krankheiten des Muskel-Skelett-Systems und des Bindegewebes repräsentiert.

Die Ursache dieser Zunahme war die Insertion einer fünften Stelle an der Kodierung, um die Abbildung im German Diagnosis Related Groups (G-DRG)-System zu ermöglichen [12].

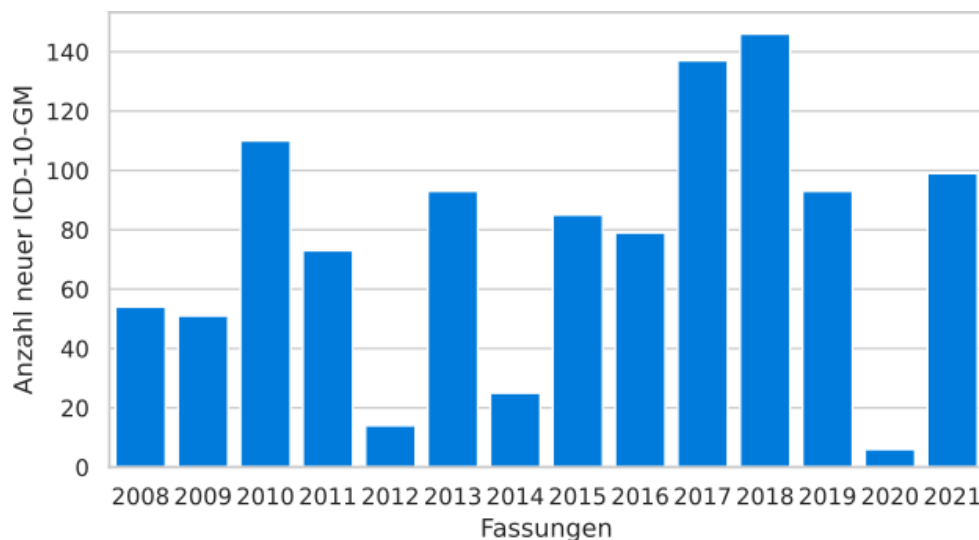


Abbildung 3.2: Anzahl neuer ICD-10-GM zwischen den Jahren 2008 und 2021

Ein wichtiger und aktueller Punkt sind die meldepflichtigen Krankheiten im Laufe der Jahre. Dieses Phänomen ist in der Abbildung 3.4 dargestellt. Es ist zu erkennen, dass neue meldepflichtige Codes in den Jahren 2010, 2016 und 2021 definiert wurden (Tabelle 3.3). Einige Ursachen davon sind Pandemien und Epidemien wie die Influenza Varianten zwischen 2009 und 2010 [13, 14], und jetzt aktuell seit Februar 2020 die Verbreitung des Corona Virus in Europa [15] mit deren gesundheitlichen Folgen [16]. Andere Gründe für neue Kodierungen sind vorgenommene Umstrukturierungen der Kodebereiche. Ein Beispiel davon sind die Bereiche für Dengue. Diese wurden von der Weltgesundheitsorganisation (WHO), 2016, umstrukturiert, und die dreistelligen Schlüsselnummern A90 und A91 wurden gestrichen und deren Inhalte in den neuen Kodebereich A97.- verlagert und auf vierstellige Subkategorien ausdifferenziert [17].

Ein interessanter Aspekt der meldepflichtigen Krankheiten in der Tabelle 3.3 ist die Diagnose B17.9 Akute Virushepatitis, nicht näher bezeichnet. Diese Krankheit war unter K72.0 Akutes und subakutes Leberversagen zugeordnet, erst 2010 wurde zur Kenntnis genommen, dass es sich bei der akuten Hepatitis in manchen Fällen um eine akute infektiöse Hepatitis handelt. Aus diesem Grund hatte die WHO die Schlüsselnummer B17.9 zur Verfügung gestellt [18].

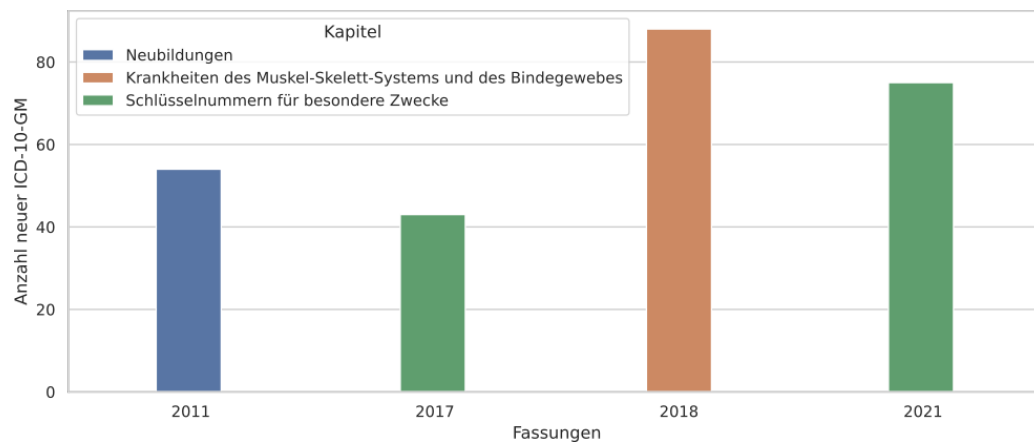


Abbildung 3.3: Kapitel mit den meisten neuen Kodierungen zwischen 2008 und 2021

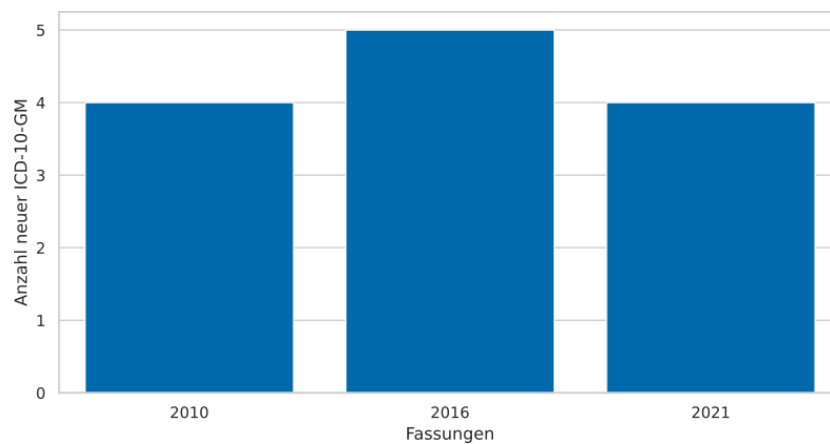


Abbildung 3.4: Menge der neuen meldepflichtigen Diagnosen zwischen den Jahren 2008 und 2021

Tabelle 3.3: Meldepflichtige Krankheiten

Fassung	ICD-10-GM	Titel
2010	B17.9	Akute Virushepatitis, nicht näher bezeichnet
2010	U69.2-!	Sekundäre Schlüsselnummern für besondere epidemiologische Zwecke
2010	U69.20!	Influenza A/H1N1 Pandemie 2009 [Schweinegrippe]
2010	U69.21!	Influenza A/H5N1 Epidemie [Vogelgrippe]
2016	A97.-	Dengue
2016	A97.0	Dengue ohne Warnzeichen
2016	A97.1	Dengue mit Warnzeichen
2016	A97.2	Schweres Dengue
2016	A97.9	Dengue, nicht näher bezeichnet
2021	U07.1!	COVID-19, Virus nachgewiesen
2021	U07.2!	COVID-19, Virus nicht nachgewiesen
2021	U10.-	Multisystemisches Entzündungssyndrom in Verbindung mit COVID-19
2021	U10.9	Multisystemisches Entzündungssyndrom in Verbindung mit COVID-19, nicht näher bezeichnet

3.4 Modifizierte ICD-10-GM

Jedes Jahr werden Änderungen in den Metadaten der ICD-10-GM vorgenommen. Solche Modifikationen sind in der Tabelle 3.4 aufgelistet. Wie in der Sektion 3.1 genannt wurde, sind die meisten Änderungen an den Spalten für die dritten, vierten und fünften Stellen der Codes, weil diese Felder erst im Jahr 2013 entstanden sind [9]. Ein weiterer Aspekt in Bezug auf die zahlreichen Änderungen ist, dass die Werte verschiedener Felder bei manchen ICD-10-GM in früheren Fassungen noch nicht definiert waren, wie in der Tabelle 3.4 in den Fällen der Mortaliätslisten verdeutlicht wird. Die Werte solcher Listen bei den meisten Codes im Jahr 2007 waren nicht definiert, und waren mit dem Wert „UNDEF“ gesetzt. Erst 2008 wurden diese Felder durch einen spezifischen Wert ersetzt. Die zahlreichen Modifikationen an den Klassentiteln sind durch Vorschläge zur Weiterentwicklung der Klassifikationen entstanden [19, 20].

Ein interessanter Aspekt an der Stelle des Geschlechtsbezugs ist, dass 52 Diagnosen mit Geschlechtsbezug der 68 an dieser Stelle geänderten Schlüsselnummer, Krankheiten der Brustdrüse sind. Diese ICD-10-GM waren früher auf das weibliche Geschlecht bezogen, jetzt ist der Geschlechtsbezug dieser Kodierungen irrelevant. Viele dieser Änderungen wurden in den Jahren 2008 und 2016 vorgenommen. An dieser Stelle können wir nicht erklären, wieso diese Art von Krankheiten nicht immer ohne

Tabelle 3.4: Liste der Metadaten und Anzahl an Änderungen

Änderungen	Metadaten
15578	Dritte Stelle
13871	Vierte Stelle
6949	Bezug zur Mortalitätsliste 3
6348	Bezug zur Mortalitätsliste 1
2	Bezug zur Mortalitätsliste 1 (definiert)
5061	Fünfte Stelle
1111	Klassentitel
308	Untere Altersgrenze
36	Untere Altersgrenze (relevant)
299	Bezug zur Morbiditätsliste
11	Bezug zur Morbiditätsliste (definiert)
286	Obere Altersgrenze
14	Obere Altersgrenze (relevant)
162	Anwendung der Laborausschlussziffer des einheitlichen Bewertungsmaßtabs (EBM)
118	Art der Vier- und Fünfsteller
69	Bezug zur Mortalitätsliste 4
68	Geschlechtsbezug
64	Bezug zur Mortalitätsliste 2
2	Bezug zur Mortalitätsliste 2 (definiert)
44	Kategorie der Gruppe
40	Arzt-Meldepflicht
31	Sehr seltene Krankheit in Mitteleuropa
6	Belegung des Kodes
3	§295 SGB V
2	Klassifikationsebene

Geschlechtsbezug waren, denn seit 1930 existieren medizinische Publikationen über Brustdrüsen Anomalien bei Männern [21]. In den heutigen Tagen gibt es eine leichte Zunahme an Männern mit Brustkrebs [22] und die erste deutschsprachige Fassung der ICD-10 für die Zwecke des SGB V wurde schon 1996 erarbeitet [23].

Tabelle 3.5: Beispiel von seltenen Krankheiten in Mitteleuropa anhand von ICD-10-GM der Bereiche A36.- Diphtherie und A80.- Akute Poliomyelitis [Spinale Kinderlähmung]. Die alten und neuen Werte sind in der Spalte „Alt“ bzw. „Neu“. „N“ Nein, „J“ Ja.

Fassung der Änderung	ICD-10-GM	Titel	Alt	Neu
2011	A36.-	Diphtherie	N	J
2011	A36.0	Rachendiphtherie	N	J
2011	A36.1	Nasenrachendiphtherie	N	J
2011	A36.2	Kehlkopfdiphtherie	N	J
2011	A36.3	Hautdiphtherie	N	J
2011	A36.8	Sonstige Diphtherie	N	J
2011	A36.9	Diphtherie, nicht näher bezeichnet	N	J
2011	A80.-	Akute Poliomyelitis [Spinale Kinderlähmung]	N	J
2011	A80.0	Akute paralytische Poliomyelitis durch Impfvirus	N	J
2011	A80.1	Akute paralytische Poliomyelitis durch importiertes Wildvirus	N	J
2011	A80.2	Akute paralytische Poliomyelitis durch einheimisches Wildvirus	N	J
2011	A80.3	Sonstige und nicht näher bezeichnete akute paralytische Poliomyelitis	N	J
2011	A80.4	Akute nichtparalytische Poliomyelitis	N	J
2011	A80.9	Akute Poliomyelitis, nicht näher bezeichnet	N	J

Von den sehr seltenen Krankheiten in Mitteleuropa sind 20 der 31 Diagnosen belegte Schlüsselnummern, also Kodierungen ohne Unterkategorien und somit im alltäglichen Gebrauch. Davon gehören 12 zu den Bereichen A36.- Diphtherie und A80.- Akute Poliomyelitis [Spinale Kinderlähmung]. Diese sind in der Tabelle 3.5 aufgelistet. Diese Diagnosen sind heutzutage selten, Ursache ist die erfolgreiche Impfkampagne in Deutschland, sodass die vom Robert-Koch-Institut (RKI) jährlich selten gemeldeten Fällen in den heutigen Tagen importiert sind [24].

3.5 Gelöschte ICD-10-GM

In jeder Fassung von 2008 bis 2021 werden ICD-10-GM gelöscht. Dieses Verhalten ist in der Abbildung 3.5 repräsentiert. Die meisten Schlüsselnummern wurden im Jahr 2013 gelöscht, da eine umfangreiche Umstrukturierung in der Systematik in diesem Jahr vorgenommen wurde. Die Kapitel, die in diesem Jahr am meisten betroffen waren, sind in der Abbildung 3.6 dargestellt. Bei der semantischen Umstrukturierung im Jahr 2013 wurden bestimmte Codebereiche im Kapitel **Krankheiten des Kreislaufsystems** umfänglich überarbeitet, viele fünfstellige Klassifizierungen wurden vom Kapitel **Krankheiten des Muskel-Skelett-Systems und des Bindegewebes** entnommen, und im Kapitel **Schlüsselnummern für besondere Zwecke** wurden zahlreiche ICD-10-GM umgebaut [25].

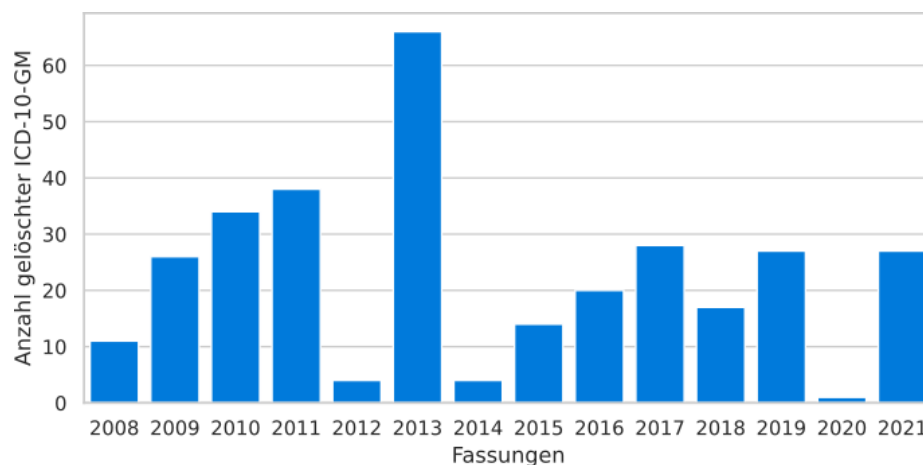


Abbildung 3.5: Anzahl gelöschter ICD-10-GM zwischen den Jahren 2008 und 2021

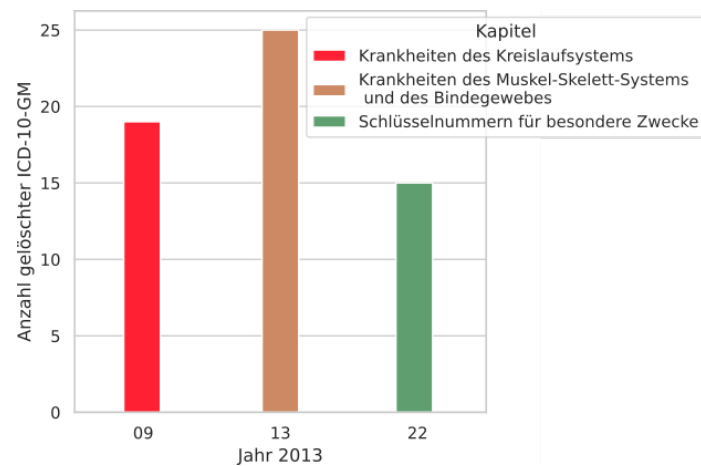


Abbildung 3.6: Meist betroffene Kapitel beim Löschen 2013

Die Abbildung 3.7 zeigt, dass die meisten Codes, nämlich 295 der 321 gelöschten Kodierungen, terminale Schlüsselnummer waren, also ICD-10-GM, die in den Texten der Dokumentation der Diagnosen benutzt werden sollen. Von diesen terminalen Kodierungen wurden 115 erweitert, wie in der Sektion 3.3 anhand des Beispiels von U81! erläutert wurde. Von den nicht terminalen Kodierungen wurden dann 11 erweitert. Ein Beispiel davon ist Z75.2- Wartezeit auf eine Untersuchung oder Behandlung. Diese nicht terminale Kodierung wurde gelöscht, und von dem terminalen Code Z75.2 mit demselben Titel ersetzt.

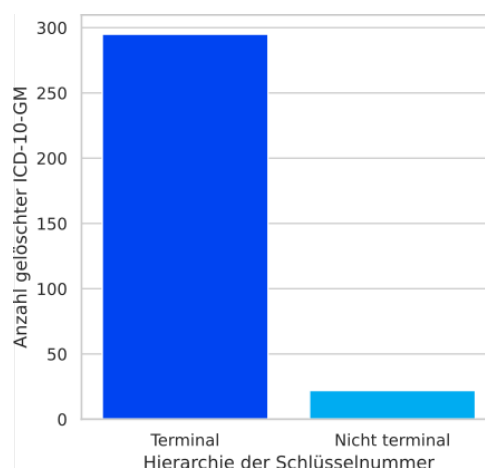


Abbildung 3.7: Hierarchie der gelöschten ICD-10-GM

3.6 Wiederverwendete ICD-10-GM

Ein unerwartetes Phänomen dieser Projektarbeit war die Erkennung von wiederverwendeten Kodierungen. Die Tabelle 3.6 zeigt die 4 wiederverwendeten Schlüsselnummern.

Tabelle 3.6: Liste der wieder benutzten Schlüsselnummern.

Gelöscht	Wieder	ICD-10-GM	Aktueller Titel
2013	2015	M21.60	Erworbener Hohlfuß [Pes cavus]
2013	2015	M21.6-	Sonstige erworbene Deformitäten des Knöchels und des Fußes
2014	2016	N90.8	Sonstige näher bezeichnete nicht-entzündliche Krankheiten der Vulva und des Perineums
2010	2019	K55.8	Sonstige Gefäßkrankheiten des Darmes

Die Schlüsselnummern M21.6- und M21.60 wurden 2013 entnommen, da im Codebereich M20-M25 zahlreiche Schlüssel zu keinen sinnvollen Kombinationen führten [25]. Andererseits wurde 2015 die Kategorie M21.6 mit den genannten Schlüsselnummern wieder aufgenommen, um eine bessere Spezifikation der Diagnosen in dem G-DRG-System zu erreichen [26].

Die Kodierung N90.8 wurde 2014 entnommen und stattdessen wurde N90.8- Sonstige näher bezeichnete nichtentzündliche Krankheiten der Vulva und des Perineums mit weiteren Subklassifikationen zur spezifischen ICD-10-GM eingeführt [20]. Zwei Jahre später, 2016, wurde der Code N90.8- und deren Unterkategorien wieder entnommen und zu einem anderen Kapitel mit neuen Kodierungen verlagert, und die Nummer N90.8 wurde wieder eingeführt [17].

Die Kodierungsschlüssel K55.8 wurde 2010 gelöscht und die Kodierung K55.8- Sonstige Gefäßkrankheiten des Darmes mit weiteren Stellen eingeführt, um einige Diagnosen des Dünndarms besser abgrenzen zu können [18]. 2019 wurde die ICD-10-GM K55.3- Angiodysplasie des Dünndarmes mit weiteren fünfstelligen Codes eingeführt und die, 2010, eingeführten Klassifikationen wurden auf den neuen Codebereich verlagert, so dass die Kodierungen von 2010 obsolet wurden, und K55.8 wieder benutzt wurde [27].

3.7 Strukturelle Änderungen

Wie in der Subsektion 2.4.2 genannt wurde, gab es auch Änderungen in der Struktur des systematischen Verzeichnisses. Im Jahr 2013 wurde die Spalte „titel“ für den Klassentitel der Tabelle „kodes“ aus dem Dreisteller-, Viersteller- und gegebenenfalls

Fünfstellertitel zusammengesetzt. Auch in diesem Jahr wurden drei neue Felder für die einzelnen Bestandteile des zusammengesetzten Klassentitels eingefügt [9]. Eine weitere Änderung in der Tabelle „kodes“, die schon seit 2015 geplant wurde, und im Jahr 2018 vorgenommen wurde, war die Entnahme der Felder der Altersklassenformate für die untere und obere Grenze des Altenbezuges mit Angabe von Tagen, Wochen, Monaten und Jahren; und damit blieben nur die Felder mit dem Format in Tagen und Jahren, nämlich „altunt“ für die untere Grenze und „altob“ für die obere Grenze des Altenbezuges [10].

4 Schlussfolgerung

Mit der Durchführung dieses Projekts konnten wir die Besonderheiten eines Lebenszyklus der historischen ICD-10-GM Fassungen von 2007 bis 2021 darstellen. Damit wurde ein DB-Schema zur Speicherung der Information des semantischen Verzeichnisses aus dem BfArM implementiert, eine ETL-Strecke zum Import der Information in der DB wurde programmiert und eingesetzt. Am Ende der Implementierung konnten die Besonderheiten der ICD-10-GM in dem vorher genannten Zeitraum, nämlich Insertion, Modifikation, Löschung und Wiederverwendung mit SQL und Python analysiert werden. Damit wurden verschiedene Ursachen für die Änderungen in den Fassungen erkannt und detailliert analysiert.

4.1 Weitere Nutzungen des Systems

Das DB-Schema unserer Implementierung ist auch nützlich für Plausibilitätsabfragen, statistische Gruppierungen und Qualitätsmanagement in klinischen Systemen, so dass Übereinstimmung und Genauigkeit der Kodierungen zum Beispiel in mehrjährigen Übersichten nachvollziehbar ist. Noch dazu kann Mithilfe unserer Implementierung im Gesundheitswesen und in der Forschung die Dienstleistungserbringungen überwacht, Gesundheitstrends erkannt, Dienstleistungen entsprechend geplant und Metaanalysen durchgeführt werden.

Literatur

- [1] *ICD-10-GM*. Zugriff am 17-10-2021. Bundesministeriums für Gesundheit (BMG), 2021. <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/>.
- [2] *Alphabetisches Verzeichnis zur ICD-10-GM*. Zugriff am 17-10-2021. Bundesministeriums für Gesundheit (BMG). <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/alphabet/>.
- [3] *Systematisches Verzeichnis der ICD-10-GM*. Zugriff am 17-10-2021. Bundesministeriums für Gesundheit (BMG). <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/systematik/>.
- [4] *Aufbau der Vierstelligen ausführlichen Systematik der ICD-10-GM*. Zugriff am 17-10-2021. Bundesministeriums für Gesundheit (BMG). <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/systematik/systematik/>.
- [5] *Kategorie und Kode in der ICD-10-GM*. Zugriff am 25-10-2021. Bundesministeriums für Gesundheit (BMG). <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/systematik/kodestruktur/>.
- [6] *PostgreSQL: The World's Most Advanced Open Source Relational Database*. Zugriff am 17-10-2021. 2021. <https://www.dimdi.de/dynamic/de/klassifikationen/icd/icd-10-gm/>.
- [7] *LIESMICHDATEI*. Bundesministeriums für Gesundheit (BMG), Nov. 2020.
- [8] Evi Nemeth, Garth Snyder, Trent R. Hein und Ben Whaley. *UNIX and Linux System Administration Handbook*. 4. Aufl. Prentice Hall, 2010. ISBN: 978-0-13-148005-6.
- [9] *LIESMICHDATEI*. Bundesministeriums für Gesundheit (BMG), Sep. 2012.
- [10] *LIESMICHDATEI*. Bundesministeriums für Gesundheit (BMG), Sep. 2016.
- [11] *Kommentar zur ICD-10-GM Version 2017*. Bundesministeriums für Gesundheit (BMG), Sep. 2016.
- [12] *Kommentar zur ICD-10-GM Version 2018*. Bundesministeriums für Gesundheit (BMG), Sep. 2017.

- [13] A Amato-Gauci, P Zucs, R Snacken, B Ciancio, V Lopez, E Broberg, P Penttinen, A Nicoll und collective on behalf of the European Influenza Surveillance Network (EISN). „Surveillance trends of the 2009 influenza A(H1N1) pandemic in Europe“. In: *Eurosurveillance* 16.26 (2011). DOI: <https://doi.org/10.2807/ese.16.26.19903-en>.
- [14] Robyn Martin, Alexandra Conseil, Abie Longstaff, Jimmy Kodo, Joachim Siegert, Anne-Marie Duguet, Paula Lobato de Faria, George Haringhuizen, Jaime Espin und Richard Coker. „Pandemic influenza control in Europe and the constraints resulting from incoherent public health laws“. In: *BMC Public Health* 10.1 (2010). ISSN: 1471-2458. DOI: [10.1186/1471-2458-10-532](https://doi.org/10.1186/1471-2458-10-532).
- [15] Feng He, Yu Deng und Weina Li. „Coronavirus disease 2019: What we know?“. In: *Journal of Medical Virology* 92.7 (2020). DOI: <https://doi.org/10.1002/jmv.25766>.
- [16] Abdul Mannan Baig. „Chronic COVID syndrome: Need for an appropriate medical terminology for long-COVID and COVID long-haulers“. In: *Journal of Medical Virology* 93.5 (2021). DOI: <https://doi.org/10.1002/jmv.26624>.
- [17] *Kommentar zur ICD-10-GM Version 2016*. Bundesministeriums für Gesundheit (BMG), Sep. 2015.
- [18] *Kommentar zur ICD-10-GM Version 2010*. Bundesministeriums für Gesundheit (BMG), Okt. 2009.
- [19] *Kommentar zur ICD-10-GM Version 2009*. Bundesministeriums für Gesundheit (BMG), Sep. 2008.
- [20] *Kommentar zur ICD-10-GM Version 2014*. Bundesministeriums für Gesundheit (BMG), Sep. 2013.
- [21] Mohd. Ajmal Husain. „Scirrhus Cancer of the Breast in the Male“. In: *The Indian Medical Gazette* 65.6 (1930), S. 330. ISSN: 0019-5863. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5157223/> (besucht am 09.09.2021).
- [22] Sharon H. Giordano, Deborah S. Cohen, Aman U. Buzdar, George Perkins und Gabriel N. Hortobagyi. „Breast carcinoma in men“. In: *Cancer* 101.1 (2004), S. 51–57. ISSN: 1097-0142. DOI: [10.1002/cncr.20312](https://doi.org/10.1002/cncr.20312). (Besucht am 20.10.2021).
- [23] Bernd Graubner. „ICD und OPS“. In: *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 50.7 (Juli 2007), S. 932–943. ISSN: 1437-1588. DOI: [10.1007/s00103-007-0283-x](https://doi.org/10.1007/s00103-007-0283-x). <https://doi.org/10.1007/s00103-007-0283-x> (besucht am 05.10.2021).

- [24] Miriam Wiese-Posselt, Christine Tertilt und Fred Zepp. „Vaccination Recommendations for Germany“. In: *Dtsch Arztebl International* 108.45 (2011). DOI: 10.3238/arztebl.2011.0771.
- [25] *Kommentar zur ICD-10-GM Version 2013*. Bundesministeriums für Gesundheit (BMG), Sep. 2012.
- [26] *Kommentar zur ICD-10-GM Version 2015*. Bundesministeriums für Gesundheit (BMG), Sep. 2014.
- [27] *Kommentar zur ICD-10-GM Version 2019*. Bundesministeriums für Gesundheit (BMG), Sep. 2018.