



UNIVERSITÄT ZU LÜBECK  
INSTITUT FÜR  
THEORETISCHE INFORMATIK

## **Abbilden und Nutzen von Versionsübergreifenden Medizinischen Klassifikationen mittels FHIR ConceptMaps**

*Visualization and Mapping of Medical Classifications across Multiple Versions with FHIR ConceptMaps*

### **Masterarbeit**

verfasst am  
**Institut für Medizinische Informatik**

im Rahmen des Studiengangs  
**Medizinische Informatik**  
der Universität zu Lübeck

vorgelegt von  
**Simon Müller**

ausgegeben und betreut von  
**Prof. Dr. Josef Ingenerf**

mit Unterstützung von  
**Tessa Ohlsen, M. Sc.**

Lübeck, den 1. November 2024

### **Eidesstattliche Erklärung**

*Ich erkläre hiermit an Eides statt, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.*

---

Simon Müller

## **Zusammenfassung**

## **Abstract**

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Grundbegriffe	1
1.2	Verwandte Arbeiten	6
1.3	Aufbau & Beiträge dieser Arbeit	8
<b>2</b>	<b>BfArM-Daten</b>	<b>9</b>
2.1	Kode- und Umsteiger-Dateien	9
2.2	Datenintegrationsprozess	14
2.3	Datenvorverarbeitung	16
<b>3</b>	<b>Umsteiger-Suche</b>	<b>21</b>
3.1	Allgemeine Funktionen	21
3.2	Transitive Hülle	25
3.3	Vergleich der Suchalgorithmen	37
3.4	Schreiben der FHIR ConceptMap	38
<b>4</b>	<b>Web-Applikation</b>	<b>39</b>
4.1	Überblick	39
4.2	Aufbau des bfarmer -Projekts	41
4.3	Backend	43
4.4	Frontend	44
4.5	AJAX	44
<b>5</b>	<b>Zusammenfassung</b>	<b>46</b>
5.1	Ergebnisse	46
5.2	Zusammenfassung	47
5.3	Ausblick	47
	<b>Literatur</b>	<b>50</b>
<b>A</b>	<b>Appendix</b>	<b>55</b>
A.1	Abweichungen zwischen ICD-10-GM und OPS Versionen	55
A.2	Beispielergbnis der Horizontalen Suche	61
A.3	Beispielergbnis der Vertikalen Suche / ConceptMap	63

# 1

## Einleitung

Das Bundesinstitut für Arzneimittel und Medizinprodukte, kurz: BfArM, veröffentlicht jedes Jahr aktualisierte Versionen medizinischer Klassifikationen, wie zum Beispiel der Kodiersysteme ICD-10-GM und OPS. Diese Änderungen in den Codes erschweren die versionsübergreifende Analyse klinischer Daten, wie sie beispielsweise vom Deutsches Forschungsdatenportal für Gesundheit (Medizininformatik-Initiative, o. D.) zur Verfügung gestellt werden. Ein technischer Ansatz dieses Code-Mapping-Problem zu lösen wäre die Verwendung von *ConceptMaps* aus dem FIHR Standards für den Austausch elektronischer Gesundheitsdaten.

### 1.1 Grundbegriffe

#### Klassifikationen ICD-10-GM und OPS

Diese Arbeit bezieht sich auf die Systematischen Verzeichnisse von ICD-10-GM und OPS, das heißt der nach dem Code strukturierten Veröffentlichungen dieser Kodiersysteme.

#### ICD-10-GM

Zitiert aus (Gaus, 2005, Seite 97):

Die International Statistical Classification of Diseases (ICD) geht auf das Jahr 1855 zurück [...] und wird heute von der World Health Organisation (WHO) betreut. Die 10. Revision trat am 01.01.1993 in Kraft unter der Bezeichnung [...] ICD-10. Die deutschen Ausgaben werden vom Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) als dem WHO Kooperationszentrum erarbeitet. Die am 01.01.2005 in Kraft getretene Ausgabe heißt [...] ICD-10-GM 2005, wobei GM für German Modification steht.

Seit Mai 2020 ist das DIMDI in das BfArM eingegliedert. (BfArM, o. D.[c])

## 1 Einleitung

Wiederum (Gaus, 2005, Seite 98ff):

Die ICD-10-GM 2005 ist ein Ordnungssystem für *Krankheiten (Diagnosen)*, das nach dem *Ordnungsprinzip Klassifikation* aufgebaut ist und insgesamt ca. 64 000 Klassen hat. Die *Notation* ist vier- oder fünfstellig. Sie besteht aus einem Buchstaben, einer zweistelligen Zahl, einem Punkt (dient nur der Strukturierung, liefert keine Information und wird deshalb bei der Stellenzahl nicht mitgezählt) und dann noch eine (vierstellige Notation) oder zwei (fünfstellige Notation) einstellige Zahlen. [...]

Die ICD-10-GM 2005 ist hierarchisch geordnet. Bei der Notation bezeichnen die 3 Stellen vor dem Punkt ein hierarchisches Niveau, das Raum für 2 500 gleichgeordnete Klassen bietet und nur durch Überschriften weiter gegliedert ist. Die Stellen nach dem Punkt geben weitere Niveaus an, d.h. die vierstellige Notation beschreibt 2, die fünfstellige Notation 3 hierarchische Niveaus. [...]

Ein allgemeines Problem der systematischen Anordnung von Diagnosen ist, ob eine Diagnose unter dem Ort ihrer Manifestation oder unter dem ihr zugrunde liegenden Krankheitsprozess eingeordnet werden soll. *Beispiel*: Soll die Lungenentzündung unter der Lokalisation Lunge oder unter dem Krankheitsprozess Entzündung eingeordnet werden? Werden alle Krankheiten eines bestimmten Organs nebeneinander gestellt, so folgt die Systematik dem topologisch-organspezifischen Aspekt (*Topologie* = Lehre von der Lage und Anordnung der Dinge im Raum). In einer Systematik können jedoch auch alle Krankheiten mit dem gleichen Krankheitsprozess, z.B. alle Entzündungen, alle Autoimmunkrankheiten oder alle bösartigen Neubildungen, nebeneinander gestellt werden. Diese Einteilung nennt man ätiologisch, pathologisch oder nosologisch (*Ätiologie* = die Krankheit auslösende Ursache, *Pathologie und Nosologie* = Lehre von den Krankheiten). Dieses Problem ist mit dem Ordnungsprinzip Klassifikation nicht lösbar. Die [ICD-10-GM] benutzt im Wesentlichen den ätiologischen Aspekt.

Definition des Begriffs Klassifikation aus (Gaus, 2005, Seite 86f):

Von allen Ordnungsprinzipien ist die Klassifikation das einfachste. Es beruht auf dem Grundsatz: „Jedes Ding (jeder Sachverhalt) an seinen Platz“. Das zu dokumentierende Sachgebiet wird in einzelne *getrennte Sachverhalte* eingeteilt, die man als *Klassen* bezeichnet. [...] Die Klassen sind *disjunkt*, d.h. sie schließen sich gegenseitig aus und überlappen sich nicht. Jede Klasse wird durch einen Deskriptor repräsentiert. Die Klassen einer Klassifikation sind gleichzeitig *Äquivalenzklassen* von Begriffen. Im strengen Fall ist die Zuordnung einer Dokumentationseinheit zu einer Klasse eindeutig, d.h. eine Dokumentationseinheit wird genau einer Klasse zugeteilt. Eine Klassifikation ist einfach und praktisch. Sie ist sozusagen das „*natürliche Ordnungsprinzip*“. [...]

Ein *Klassifikationssystem* – ein Ordnungssystem, das nach dem Ordnungsprinzip Klassifikation aufgebaut ist – muss vollständig sein. Vollständig sein bedeutet, dass die Klassen alle Sachverhalte des dokumentarisch zu bearbeitenden Sachgebiets umfassen. [...] Durch Schaffung einer Klasse „*sonstiges*“ oder besser durch die Schaffung mehrerer Klassen mit dem Zusatz „*sonstiges*“ wird die geforderte Vollständigkeit des Klassifikationssystems [...] formal erreicht. [...] Um auf die einzelnen Klassen bequem und sicher zugreifen zu können, werden sie *hierarchisch* oder anderweitig *systematisch* angeordnet.

Vom BfArM werden Notation und Begriff bezeichnet als Kode und Titel.

### OPS

Ebenfalls zitiert aus (Gaus, 2005, Seite 101ff):

In der Medizin werden nicht nur Krankheiten, sondern auch ärztliche Tätigkeiten dokumentarisch erfasst und – ebenso wie die Diagnosen für die Abrechnung und für wissenschaftliche Zwecke verwendet. Die von der WHO erstmals 1978 herausgegebene International Classification of Procedures in Medicine (ICPM) wurde (ebenso wie die ICD-10) vom DIMDI ins Deutsche übertragen und an deutsche Verhältnisse angepasst. Die am 01.01.2005 in Kraft getretene Ausgabe heißt „Operationen- und Prozedurenschlüssel – Internationale Klassifikation der Prozeduren in der Medizin (OPS 2005)“. Der OPS 2005 ist – wie die Bezeichnung „Schlüssel“ ausdrückt – eine Klassifikation mit numerischer Notation. Allerdings haben in einigen Bereichen die 10 gleichgeordneten Klassen nicht ausgereicht, deshalb treten an der 5. und 6. Stelle der Notation gelegentlich Buchstaben auf. Von den insgesamt 14 000 Klassen betreffen etwa 70% Operationen, ca. 15% nichtoperative therapeutische Maßnahmen, 8% diagnostische Maßnahmen, 4% bildgebende Diagnostik und 2% ergänzende Maßnahmen.

Die *Notation* des OPS 2005 ist vier- bis sechsstellig. Die erste Stelle enthält eine 1, 3, 5, 8 oder 9 und bezeichnet einen der eben genannten Bereiche (z.B. 1 = diagnostische Maßnahmen). Es folgt ein Bindestrich und eine dreistellige Zahl. Je nach Detaillierungsgrad ist damit die Notation beendet oder es folgen noch ein Punkt und weitere 1 bis 2 Stellen. [...] In der fünften und sechsten Stelle der Notation (d.h. nach dem Punkt) können Ziffern durch die Buchstaben x oder y ersetzt werden, es bedeutet x = sonstige und y = nicht näher bezeichnet.

### FHIR ConceptMap

FHIR steht für „Fast Healthcare Interoperability Resources“.

### HL7 FHIR

Zitiert aus (Heckmann, 2022, Seite 309):

„Health Level 7“ wurde 1987 gegründet, um Standards für klinische Informationssysteme zu erarbeiten. [...] HL7 ist vom nationalen amerikanischen Normeninstitut (ANSI) seit 1994 akkreditiert. HL7-verbundene Organisationen existieren inzwischen in über 40 Ländern. Die erste nationale Partnergesellschaft wurde 1993 in Deutschland gegründet. Die Aufgabe dieser sogenannten „Affiliates“ ist es, die Verbreitung und Implementierung der internationalen HL7-Standards in dem jeweiligen Land zu fördern und gegebenenfalls erforderliche Anpassungen und Erweiterungen für deren Nutzung zu entwickeln. FHIR ist die dritte Generation von Interoperabilitätsstandards aus der Feder von HL7.

Die Entwicklung begann im Jahre 2011 als Reaktion auf die Forderungen aus der Industrie nach einer standardisierten Lösung für die Entwicklung webbasierter Applikationen für das Gesundheitswesen. [...] Die vierte, 2018 veröffentlichte Version „R4“ enthält erstmals normative Inhalte.

Ebenfalls aus (Heckmann, 2022, Seite 310):

Im Gegensatz zu älteren HL7-Standards [...] bedient sich FHIR erstmals moderner web-basierter Technologien, die auch jenseits des Gesundheitswesens weit verbreitet und bei Entwicklern bekannt und beliebt sind. FHIR legt den Fokus auf die query-getriebene Art der Kommunikation, die dazu dient, Anwendungen und Anwendern die benötigten Daten zum benötigten Zeitpunkt in adäquatem Umfang und Format zur Verfügung zu stellen. Dazu bedient sich FHIR dem Representational State Transfer (kurz: „REST“), einem Paradigma, das die Struktur und das Verhalten des World Wide Webs abstrahiert und basierend auf denselben Technologien und Grundlagen eine einheitliche Form der Kommunikation von verteilten Systemen und Webservices definiert. Die auszutauschenden Datenobjekte werden als „Ressourcen“ bezeichnet, auf denen jeweils sogenannte „CRUD“-Interaktionen (create, read, update, delete) über die HTTP-Methoden GET, PUT, POST und DELETE ausgeführt werden können.

Die Autorin hebt außerdem folgende Eigenschaften hervor:

- „FHIR als Baukasten“: Standards und Leitfäden in FHIR sind selbst Ressourcen. Sie können erweitert und für verschiedene Zwecke eingesetzt werden.
- „FHIR als Community-Standard“: Die FHIR-Spezifikation ist frei verfügbar und wird von einer Community entwickelt, statt von einem geschlossenen Gremium.

Die am meisten verbreitetsten Implementationen von FHIR sind die als freie Software lizenzierte API-Server *HAPI*, sowie der proprietäre Terminologie-Server *Ontoserver*, welcher unter anderem im australischen Gesundheitssystem eingesetzt wird. Für mehr dazu siehe (Braunstein, 2022).

### REST

Representational State Transfer ist –wie oben erwähnt– ein Paradigma für die Software-architektur von verteilten System, welches erstmals in (Fielding, 2000) dargelegt wurde. Es besteht aus sechs Einschränkungen:

1. Kommunikation per Client-Server-Modell
2. Zustandslosigkeit: Jede Nachricht zwischen Client und Server muss alle Informationen enthalten, die zum Verständnis der Nachricht benötigt werden.
3. Erhöhung der Effizienz durch Einsatz von Caching
4. Einheitliche Schnittstellen:
  - 4.1 Jede Ressource wird durch eine URL identifiziert
  - 4.2 Ressourcen werden verändert durch Repräsentationen, zum Beispiel strukturierte Dateiformate
  - 4.3 Die Inhalte der Nachrichten sind selbstbeschreibend
  - 4.4 Informationen werden vom Anwendungsserver dynamisch bereitgestellt über Hypermedien wie HTTP

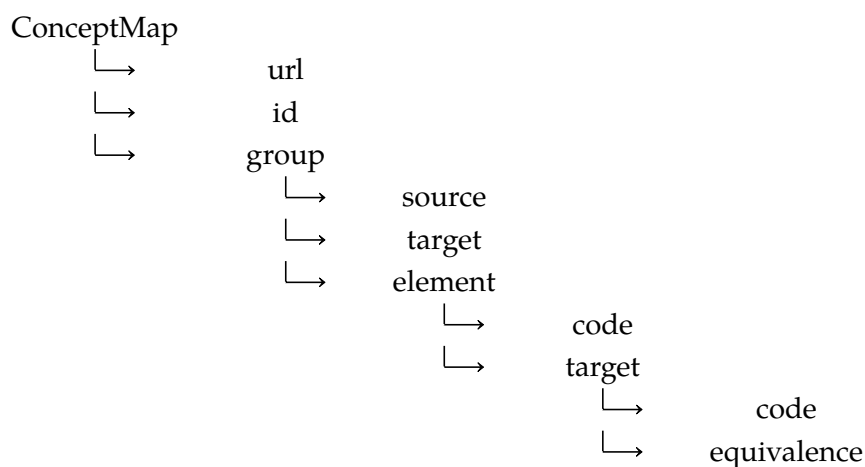


5. Mehrschichtigkeit: Komplexität wird hinter den offenen Endpunkten verborgen
6. „Code on Demand“: Es soll die Möglichkeit für Clients bestehen, neue Medientypen verstehen zu lernen

## ConceptMap

Die ConceptMap ist eine FHIR Ressource, die zum Mapping zwischen Kodiersystemen vorgesehen ist. Die Definition ist frei verfügbar, zum Beispiel für FHIR Release 4: (HL7, 2019, FHIR R4).

Die Komponenten, die für ein Mapping zwischen Versionen der ICD-10-GM und des OPS mindestens benötigt werden, sind folgende:



**Abbildung 1.1:** Hierarchischer Aufbau der FHIR Ressource „ConceptMap“.

- *url*: Eine absolute URI (Uniform Resource Identifier), welche einmal pro ConceptMap enthalten ist und diese global eindeutig identifiziert. Idealerweise handelt es sich um eine existierende URL. Laut Spezifikation ist die Angabe eigentlich optional, aber sie wird von HAPI vorausgesetzt.
- *id*: Die logische ID der Ressource; das heißt sie ist innerhalb eines Servers eindeutig. Sie kann nicht verändert werden und ist wie die URL pro ConceptMap einmalig eingetragen, sowie optional. Aber es ist natürlich sinnvoll die ID dafür zu verwenden, Zweck und Umfang der ConceptMap zu kennzeichnen.
- *group*: Eine Gruppe an Mappings, die alle das selbe Start- und Zielsystem haben. Eine ConceptMap kann mehrere Gruppen enthalten.
- *source*: Eine absolute URI, die das Start- beziehungsweise Ursprungssystem der Konzepte für das Mapping identifiziert. Sie kann einmal pro *group* angegeben werden. Bei einer in sich geschlossenen ConceptMap, das heißt ohne Referenz auf eine FHIR „Code System“ Ressource, kann dieser Eintrag auch frei gewählt sein.
- *target*: Wie *source* nur für das Zielsystem.

- *element*: Mappings für ein einzelnes Konzept der Ursprungsversion auf ein oder mehrere Konzepte der Zielversion. Pro *group* sind mehrere *element*-Einträge möglich.
- *code*: Ein Eintrag pro *element*, der dieses identifiziert.
- *target*: Konzept(e) der Zielversion, für das Mapping von *element*.
- *code*: Kode, der das Zielkonzept identifiziert.
- *equivalence*: Die Relation zwischen Konzepten der Ursprungs- und Zielversion. In FHIR R5 heißt diese Komponente *relationship*.

Wie die Relationen angegeben werden, unterscheidet sich zwischen FHIR Release 4 und 5 (HL7, 2023). Release 6 wird hierbei unverändert zu R5 bleiben. Für die Überleitungen zwischen den Versionen der ICD-10-GM und des OPS gibt es vier Fälle:

1. Es gibt keine Veränderung. R4 und R5: *equivalent*.
2. Der Kode wurde verändert. R4: *relatedto*. R5: *related-to*.
3. Die Überleitung erfolgt auf mehrere mögliche Kodes. R4: *wider*.  
R5: *source-is-narrower-than-target*.
4. Der Kode wurde entfernt. R4: *unmatched*.  
In R5 ist kein *target*-Eintrag mehr notwendig, sondern es wird stattdessen auf der gleichen Ebene wie der *code* eine Komponente *noMap* mit Wert *true* eingetragen.

Zusätzlich gibt es noch die Komponente *unmapped* pro *group*. Hier kann angegeben werden, wie Konzepte behandelt werden sollen, die zwar in dem unter *source* referenzierten Kodiersystem enthalten sind, für die es allerdings keinen Eintrag im Mapping als *element*→*code* gibt. Da bei den Überleitungen zwischen den Versionen der ICD-10-GM und des OPS meistens die große Mehrheit der Kodes unverändert bleibt, wäre es also sinnvoll, um die Größe der ConceptMap gering zu halten, alle Kodes mit *equivalent*-Relation gar nicht einzutragen und stattdessen den *unmapped*-Modus „provided“ zu verwenden, der diese Kodes einfach wie angegeben als „Ziel“ des Mappings zurückgibt, was ebenfalls der Äquivalenz entspricht. Leider wird aber *unmapped* in der aktuellen HAPI-Version nicht unterstützt und von Ontoserver nur im beschränkten Maße: (Lawley, 2023).

## 1.2 Verwandte Arbeiten

Mappingqualität

Medicats (ICD-10-GM Versionen)

Beispiele für FHIR-Mappings

UMLS

Oncology -> SNOMED CT

SNOMED als Interlingua

## Medicats

Medical Classification and Terminology Systems in a Secondary Use Context: Challenges and Perils (Hund u. a., 2016)

Medicats (Hund, 2018)

## ConceptMap für UMLS

Building Interoperable FHIR-based Vocabulary Mapping Services: A Case Study of OH-DSI Vocabularies and Mappings

using multiple groups (Saripalle, 2019)

instead? – An Interoperable UMLS Terminology Service Using FHIR

## Interlingua SNOMED CT

SNOMED (Philipp u. a., 2022)

Probleme:

- Uneindeutig, med. Fachpersonal benötigt
- Deutsche Version SNOMED CT
- OPS
- Wenn Mapping von ICD-10-GM auf SNOMED CT existiert, kann versionsübergreifendes Mapping für ICD-10-GM immer noch hilfreich sein.

Kode	Titel
203533003	Peroneal spastic flat foot
90374001	Acquired spastic flat foot
53226007	Talipes planus
44480001	Acquired flexible flat foot
203531001	Hypermobile flat foot
203534009	Acquired pes planus
203532008	Rigid flat foot

## Qualitätsstandards für Mappings

Es folgt eine Zusammenfassung verschiedener Grundprinzipien und Maßstäbe für das Erstellen von Mappings im Gesundheitswesen. In eckigen Klammern sind die Listennummern in den jeweiligen Quellen referenziert:

A „Classifications and Terminology Mapping – Principles and Best Practice“, World Health Organization, *Arbeitsgruppe: Family of International Classifications*. (WHO-FIC, 2021)

B „Principles of Mapping between Terminological Systems“, International Organization for Standardization, *Komitee: Health Informatics*. (ISO, 2014)

C „Terminology Resource Map Quality Measures“, International Organization for Standardization, *Komitee: Health Informatics*. (ISO, 2019)

1. Einsatzszenarien müssen vor Entwicklung des Mappings etabliert sein. Zweck, Umfang und die Richtung müssen klar definiert sein. Mappings sollten jeweils eine Richtung und einen Zweck haben. [A1, A2, A3, B1, B2, B9]
2. Die Urheber des Kodiersystem sollten in der Erstellung und Validierung des Mappings beteiligt sein. Mit dem Kodiersystemen verbundenen Konventionen und Richtlinien sollten eingehalten werden. [A5, B6, B7]
3. Das Team sollte nach den vorhandenen Kompetenzen strukturiert sein. Es ist förderlich, wenn ein Interesse externer Organisationen besteht. [A6, B5, B20, C9, C15]
4. Dokumentationen über Zweck, Umfang, Einschränkungen und Implementierung des Mappings müssen erstellt werden. Bei automatisierten, beziehungsweise maschinell erstellten Mappings müssen Informationen über die verwendeten Technologien und Algorithmen bereitgestellt werden. Es ist wichtig die Ergebnisse manuell zu überprüfen. [A4, A11, B8, B12, B19, C7, C10, C14]
5. Qualitätssicherung und Validierung sind essentiell; Testprotokollen müssen existieren und zukünftige Anwender des Mappings sollten involviert sein. [A7, B4, B14, B15, C11, C12]
6. Die Veröffentlichung des Mappings muss Informationen über Versionen, Update-Zyklen und Lizenzen enthalten. Ein Wartungsprozess muss definiert sein. Die Weiterentwicklung sollte durch Feedback von Anwendern gewährleistet sein und über einen Konsensbildung erfolgen. Mappings sollten möglichst aktuell sein. [A8, A9, B13, B16, B17, B18, C8, C13, C16, C17]
7. Kardinalität, Relationen, sowie Ursprung und Ziel des Mappings müssen klar definiert sein. [A12, A13, B10, B11, C4, C5, C6]
8. Mappings sollten maschinenlesbar sein. [A14, B3]

Für diese Arbeit nicht relevant sind Punkte zu manuell erstellten Mappings [A10], spezifischen Kodiersystemen [A15], Eigenschaften von Ursprung- und Zielsystem, falls diese abweichen [C1, C2], Übersetzung natürlicher Sprachen [C3].

### 1.3 Aufbau & Beiträge dieser Arbeit

abc

# 2

## BfArM-Daten

Dieses Kapitel behandelt die elektronische Verarbeitung der BfArM-Daten.

### 2.1 Kode- und Umsteiger-Dateien

Das BfArM stellt für jede neue ICD-10-GM und OPS Version Dateien für die Überleitung auf die Vorgänger-Version zum Download zur Verfügung, gebündelt jeweils in einer Zip-Datei: (BfArM, o. D.[b], Downloads).

Es handelt sich hierbei um CSV-formatierte Text-Dateien. „Comma-Separated Values“ ist ein sehr einfaches Format, um Daten zu strukturieren. Es wird ein Satzzeichen verwendet, um den restlichen Text in Spalten zu trennen – laut dem Namen normalerweise ein Komma, aber für die BfArM-Dateien wurde Strichpunkt als Trennzeichen gewählt, wahrscheinlich weil die Klassentitel auch Kommata enthalten können. Weitere Informationen zum CSV Datei-Format finden sich in (Bonney u. a., 2024, Seite 131f).

Vom BfArM werden Codes als Schlüsselnummern bezeichnet, wenn diese eindeutig sind und einzelne Überleitungen zwischen Codes werden Umsteiger genannt.

#### Kodes / Schlüsselnummern

Hier beispielhaft die ersten sieben Zeilen der Kode-Datei aus der ICD-10-GM 2024:

```
UNDEF;Undefined
A00;Cholera
A00.0;Cholera durch Vibrio cholerae 0:1, Biovar cholerae
A00.1;Cholera durch Vibrio cholerae 0:1, Biovar eltor
A00.9;Cholera, nicht näher bezeichnet
A01;Typhus abdominalis und Paratyphus
A01.0;Typhus abdominalis
```

Anmerkungen:

- Ein Strichpunkt = zwei Spalten
  1. Kode
  2. Klassentitel
- Für OPS ist das Format der Kode-Datei identisch.
- Mit Ausnahme des UNDEF-Eintrags in der ersten Zeile ist die Datei alphabetisch nach dem Kode sortiert. UNDEF ist kein ICD-10-GM- oder OPS-Kode, sondern wird für Umsteiger verwendet, um entfernte beziehungsweise neu hinzugefügte Kodes zu kennzeichnen.
- Die Datei enthält nicht-endständige Kodes – im Beispiel oben A00 und A01. Ein Kode ist endständig, wenn er keine Subkategorie hat, siehe (BfArM, o.D.[d], Kategorie und Kode in der ICD-10-GM).

### Umsteiger / Überleitungen

Im Gegensatz zu den Kodes haben die Umsteiger für ICD-10-GM und OPS unterschiedliche Formate. Hier also zuerst zwei Ausschnitte aus der Umsteiger-Datei für die ICD-10-GM 2017 Überleitung auf Version 2016:

A00.0;A00.0;A;A
A00.1;A00.1;A;A
A00.9;A00.9;A;A
A01.0;A01.0;A;A
A01.1;A01.1;A;A
-----
U06.0;UNDEF;A;
UNDEF;Z99.0;;
Z99.0;Z99.0;A;

Anmerkungen:

- Drei Strichpunkte = vier Spalten
  1. Alter Kode (2016)
  2. Neuer Kode (2017)
  3. Wenn A: automatisch überleitbar von 2016 auf 2017, sonst kein Zeichen
  4. Wenn A: automatisch überleitbar von 2017 auf 2016, sonst kein Zeichen
- Der obere Abschnitt umfasst die fünf ersten Zeilen der Umsteiger-Datei.
- Der untere Abschnitt enthält beispielhaft zwei Umsteiger mit UNDEF. UNDEF als neuer Kode heißt der alte Kode wurde entfernt. UNDEF als alter Kode heißt der neue Kode wurde hinzugefügt. In diesem Beispiel wurde Z99.0 umbenannt.

- Die Datei ist alphabetisch nach dem alten Kode sortiert und falls dieser bei mehreren Einträgen identisch ist, anschließend nach dem neuen Kode.
- Es sind nur endständige Kodes enthalten.

Dazu im Vergleich ein einzelner Umsteiger aus dem OPS 2024, Überleitung auf Version 2023:

1-100;N;1-  
100;N;A;A

Die zusätzlichen Spalten jeweils nach den Kodes speziell für den OPS sagen aus, ob Zusatzkennzeichen notwendig sind, siehe dazu auch: (BfArM, o. D.[e], Kategorie und Kode im OPS).

### „DRY“-Prinzip

„Don’t Repeat Yourself“ ist eines der Kardinalprinzipien in der Software-Entwicklung. Obwohl der Grundsatz, Wiederholungen zu vermeiden, wahrscheinlich schon in der Programmierung angewandt wird seit es diesen Beruf gibt, wurde „DRY“ erstmals 1999 ausformuliert von (Thomas und Hunt, 2019, Seite 79ff). In der zwanzigjährigen Jubiläumsausgabe verdeutlichen die Autoren, dass es ihnen hierbei nicht nur um das Schreiben von Programmcode geht, sondern vielmehr um die Intention hinter einem Prozess. Das heißt eine Änderung der Funktionalität in einer Software-Lösung sollte möglichst *nicht* mehrere Änderungen an mehreren Stellen nach sich ziehen.

Bei der Integration der BfArM-Daten kann das „DRY“-Prinzip auf zwei Arten angewandt werden.

1. Bezogen auf Kodiersysteme: Alle Funktionen sollten unabhängig davon anwendbar sein, ob es sich um ICD-10-GM- oder OPS-Daten handelt. Auch die Aufnahme eines zusätzlichen Systems, beispielsweise ATC, sollte möglichst nur Anpassungen erfordern, die durch Abweichungen in der Integration der Daten dieses Systems notwendig sind.
2. Bezogen auf Versionen: Unabhängig von der Version sollte der Prozess der Datenintegration gleich ablaufen und das gleiche Ergebnis liefern, bezogen auf die Datenstruktur. Jede Abweichung zwischen Versionen sollte nur eine möglichst einfach zu implementierende Modifikation des Gesamtprozesses darstellen. Konkret heißt das beim Hinzufügen einer neuen Version, dass an nur einer Stelle die Abweichungen von der Standardversion angegeben werden sollten und der Datenintegrationsprozess danach einmal angestoßen wird. Idealerweise ändert sich bei einer neuen Version nur die Versionsnummer und die Download-URL.

### Standardverfahren und Abweichungen

Im diesem Abschnitt werden alle Abweichungen der ICD-10-GM- und OPS-Versionen von der als Standard gewählten Version 2024 in Tabellen aufgeführt. Konkret gemeint

sind damit: Version, Download-URL, Pfad der Kode- und Umsteiger-Dateien, Sonstiges. Diese Informationen können dem Datenintegrationsprozess in einem strukturierten Dateiformat zur Verfügung gestellt werden, siehe Abschnitt 4.1.

### Liste aller Versionen

Im Anhang A.1 befinden sich Tabellen, die alle Abweichungen zwischen den Versionen für ICD-10-GM und OPS bezogen auf 2024 enthalten.

### Datei-Adressen und -Pfade

Die Download-URL der Zip-Dateien setzt sich wie folgt zusammen:

`https://multimedia.gsb.bund.de/BfArM/downloads/klassifikationen/ ...`  
& für ICD-10-GM: `icd-10/ ...`  
& für OPS: `ops/ ...`  
& einen pro Version unterschiedlichen Teil, siehe URL-Eintrag in den Tabellen.  
Die URL dient damit als „Single Source of Truth“ (Bonney u. a., 2024, Seite 257).

Die Kode- und Umsteiger-Dateien sind in einem Verzeichnis enthalten:

Klassifikationsdateien

Wenn die Tabellen einen Verzeichnis-Eintrag enthalten, wird dieser vorangestellt.

Zum Beispiel für ICD-10-GM Version 2021:

`icd10gm2021syst-ueberl-20201111/Klassifikationsdateien`

Der Pfad der Kode-Datei lautet:

Verzeichnis `...`  
& für ICD-10-GM: `icd10gm ...`  
& für OPS: `ops ...`  
& die Version `...`  
& `syst.txt`

Also zum Beispiel: `Klassifikationsdateien/icd10gm2024syst.txt`

Wenn die Tabellen einen Codes-Eintrag enthalten, wird dieser stattdessen verwendet.

Die Umsteiger-Datei funktioniert ähnlich, nur dass der Dateiname normalerweise so ist:

für ICD-10-GM: `icd10gm ...`  
für OPS: `ops ...`  
& Version `...`  
& `syst_umsteiger_` `...`  
& Vorgänger-Version `...`  
& `_` `...`  
& Version `...`  
& `.txt` Zum Beispiel: `ops2024syst_umsteiger_2023_2024.txt`



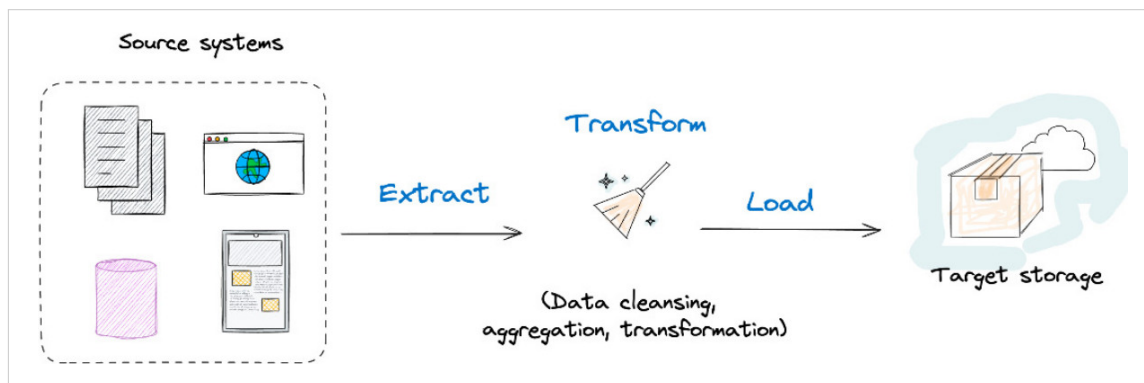
### Sonstige Abweichungen

- Vorab-Version  
Diese Version hat noch keine Seite für die Kode-Suche, siehe Abschnitt 4.5.
- Zip-Unterdatei  
Die Zip-Datei der 2022 Versionen enthielt weitere Zip-Dateien. Vorher wurden alle Dateien zu einer Versionen nach Verwendungszweck nur in Unterverzeichnisse gegliedert, weswegen die gebündelte Zip-Datei insgesamt relativ groß wurde. Ab 2023 werden die Zip-Unterdateien separat zum Download angeboten.
- ISO-8859-1  
Vor 2009 waren Dateien in ISO-8859-1 kodiert, auch Latin-1 genannt, statt UTF-8. Mehr dazu auf der nächsten Seite.
- Punkt-Strich-Notation, Kreuz-Stern-System  
Die Codes der frühesten ICD-10-GM Versionen hatten Sonderzeichen gemäß (BfArM, o. D.[d], Kategorie und Kode in der ICD-10-GM).
- 6-Spalten-Umsteiger  
Umsteiger älterer ICD-10-GM Versionen enthielten Informationen zur Mehrfachkodierung.
- Nicht endständige Umsteiger  
Im Gegensatz zu allen anderen Überleitungen sind die Umsteiger-Einträge für die ICD-10-GM 2.0 auf 1.3. auch für nicht-endständige Codes enthalten.
- None statt UNDEF  
Von OPS 2009 bis 2004 wurde statt UNDEF der Bezeichner „None“ verwendet.
- KOMBI-Kode  
OPS Versionen 2.1 und 2.0 enthalten in der Codes-Datei einen zusätzlichen Eintrag: KOMBI, „Kombinationsschlüsselnummer erforderlich“.
- 6-Spalten-Umsteiger (altes Format), 5-Spalten-Umsteiger, 4-Spalten-Umsteiger  
Die Umsteiger der OPS-Versionen von 2009 bis 2005 waren anders formatiert, weil 2005 die Informationen bezüglich Zusatzkennzeichen hinzukamen und bis 2009 die Spalten unterschiedlich angeordnet waren als in allen neueren OPS Versionen.
- 6-Spalten-Umsteiger (ursprüngliches Format)  
Die Umsteiger für den OPS 2.1 enthielten zusätzliche Spalten wegen Mehrfachverschlüsselung wie die älteren ICD-10-GM Versionen.
- 3-Spalten-Umsteiger  
OPS 2.0 zeigte mit nur einer Spalte an, ob automatische Überleitungen möglich sind.
- Keine Überleitung  
Aus der ältesten Version, die Überleitungen enthält, wird zusätzlich die Codes-Datei für die Vorgänger-Versionen verarbeitet.

## 2.2 Datenintegrationsprozess

Wie erwähnt durchlaufen alle Daten unabhängig von Version und Kodiersystem den gleichen Integrationsprozess. Dieser orientiert sich an dem klassischen „Extract-Transform-Load“ Modell, siehe (Bonnefoy u. a., 2024, Seite 247ff).

1. *Extract*: Die Daten werden in einem bestimmten Format aus einem Quell-System extrahiert.
2. *Transform*: In einem oder mehreren Prozessen werden die Daten in ein standardisiertes Format transformiert, was zum Beispiel Bereinigung, Validierung und Imputation (Generieren fehlender Daten) beinhalten kann.
3. *Insert*: Die Daten werden in ein Ziel-System integriert, um dort von weiteren Applikationen verwendet zu werden.



**Abbildung 2.1:** ETL-Modell nach (Bonnefoy u. a., 2024, Seite 63)

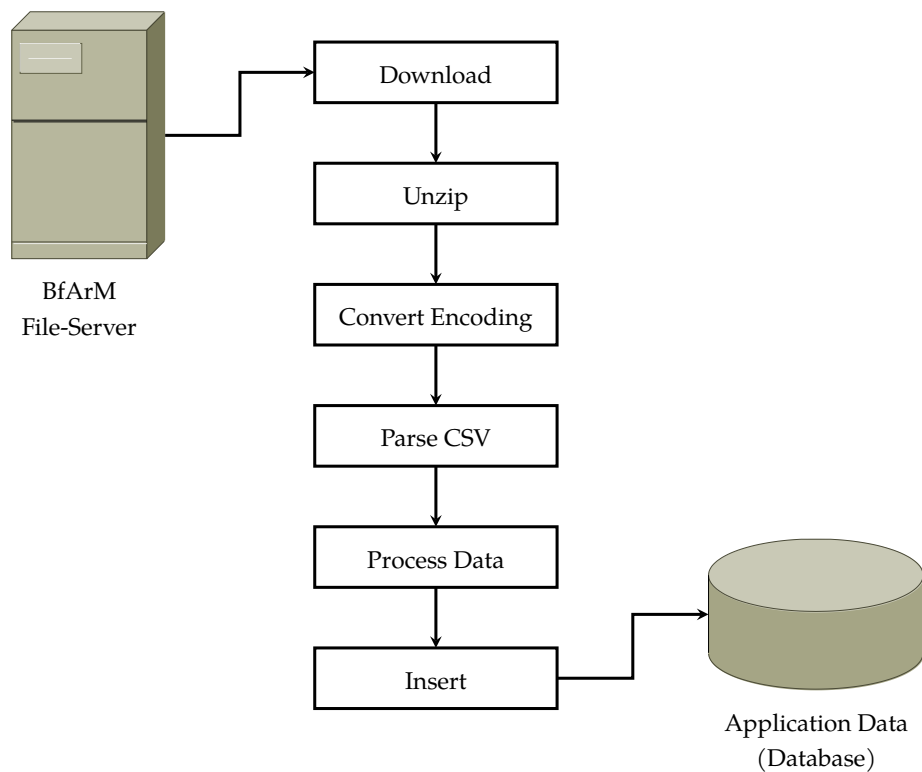
Für die BfArM-Daten sieht der Integrationsprozess konkret so aus:

1. *Download*: Die Zip-Dateien werden heruntergeladen. Alternativ kann geprüft werden, ob die Dateien schon lokal vorhanden sind mit einem bestimmten Pfad, der sich nach Kodiersystem und Versionsnummer immer gleich zusammensetzt, zum Beispiel: (Projektverzeichnis)/files/icd10gm2024.zip. Die Download-Funktion sollte die Zip-Dateien ebenfalls unter diesem Pfad abspeichern, falls so gewünscht.
2. *Unzip*: Die Kodes- und Umsteiger-Dateien werden aus der Zip-Datei extrahiert. Normalerweise muss dafür nicht das ganze Archiv in temporäre Dateien entpackt werden – außer eventuell bei den Versionen 2022, weil das Extrahieren verschachtelter Zip-Dateien eher eine Nischenanwendung ist und nicht unbedingt standardmäßig von Programmiersprachen oder Bibliotheken unterstützt wird.
3. *Convert Encoding*: Die in ISO-8859-1 kodierten Kodes-Dateien müssen in UTF-8 umgewandelt werden. In (Fernández und Manuel, 2022) werden die beiden Zeichenkodierungen genauer erklärt, aber für die BfArM-Daten ist eigentlich nur relevant, dass Umlaute mit unterschiedlichen Werten kodiert sind. Also würde das Einlesen eines in ISO-8859-1 kodierten Umlauts als UTF-8 ein anderes Zeichen als Resultat ergeben.

## 2 BfArM-Daten

Die Umsteiger-Dateien sind davon nicht betroffen, weil in diesen keine Umlaute enthalten sind.

4. *Parse CSV*: Ein Parser wandelt eine Datei in eine Datenstruktur um; für CSV sollte jede Programmiersprache so eine Funktion standardmäßig zur Verfügung stellen. Für die BfArM-Dateien ist das Ergebnis ein zweidimensionales Array mit zwei Spalten für die Codes, beziehungsweise drei bis sechs Spalten für die Umsteiger je nach Kodiersystem und Version.
5. *Process Data*: Aufgrund der oben erwähnten Abweichungen ist die Vorverarbeitung der Daten der komplexeste Schritt und wird im nächsten Abschnitt genauer erklärt. Außerdem müssen nicht alle Daten gespeichert werden. Vor allem in Bezug auf die Zip-Dateien ergibt das eine Reduktion der Datenmenge um etwa einen Faktor von zehn.
6. *Insert*: Die bearbeiteten Daten werden für die Verwendung durch Applikationen gespeichert. Zum Beispiel für eine relationale Datenbank werden pro Dateityp, Kodiersystem und Version eine Tabelle angelegt und die Daten in diese geschrieben. Konkret für SQL müssen außerdem die Hochkommata in den Codes-Dateien beachtet werden.

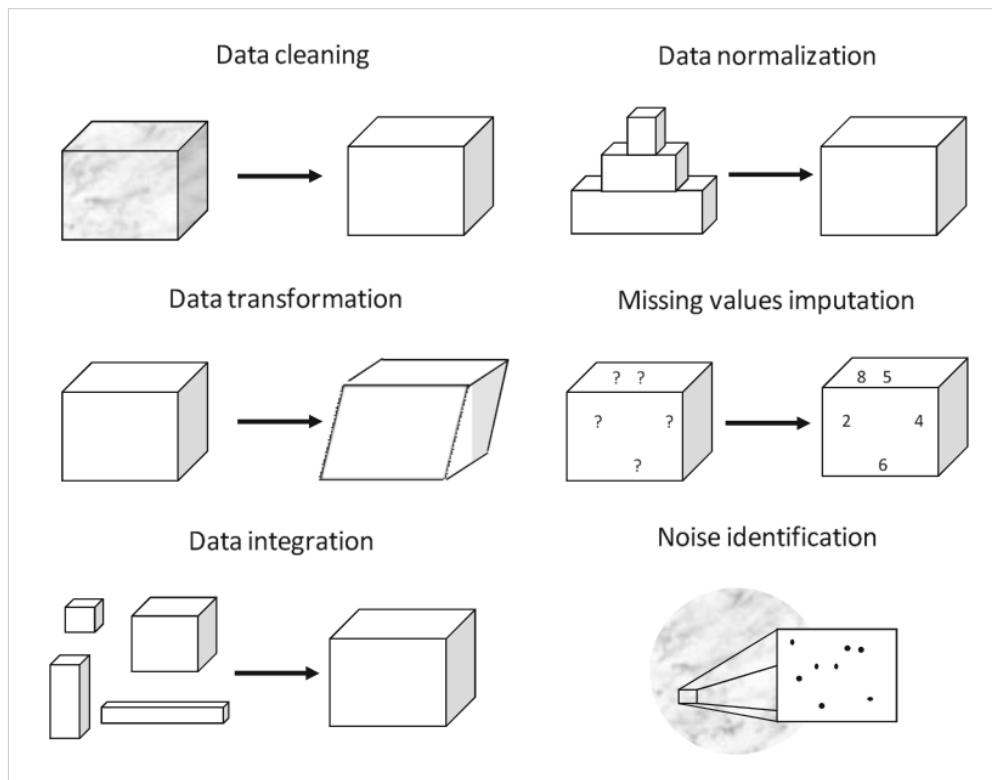


**Abbildung 2.2:** BfArM-Datenintegrationsprozess

## 2.3 Datenvorverarbeitung

„Data Preprocessing“ ist ein wichtiger Schritt in Feldern der Informatik wie *Machine Learning* und *Big Data*. In (García u. a., 2016) werden mehrere Methoden vorgestellt, wovon folgende in der Verarbeitung der BfArM-Daten zur Verwendung kommen:

1. *Data Cleaning*: Daten werden bereinigt, was sowohl das Korrigieren einzelner Werte, als auch das Entfernen überflüssiger Datensätze beinhaltet. Letzteres wird *Instance Reduction* genannt.
2. *Data Normalization*: Umwandlung der Datensätze auf ein bestimmtes Format.
3. *Data Integration*: Ein Datensatz wird durch zusätzliche Informationen bereichert, beziehungsweise mehrere Informationen werden zu einem Datensatz kombiniert.
4. *Missing values imputation*: Falls Informationen fehlen, müssen die betroffenen Datensätze mit einer bestimmten Logik behandelt werden oder alternativ können Daten durch eine Zufallsfunktion simuliert werden.



**Abbildung 2.3:** „Preprocessing Tasks“ aus (García u. a., 2016, Seite 4)

Die folgenden Unterabschnitte erklären die Vorverarbeitungsschritte für die BfArM-Daten und beziehen sich damit auf die in 2.1 genannten Abweichungen. Die Schritte erfolgen in der gelisteten Reihenfolge. Obwohl die Daten nach dem CSV-Parsing schon in einer von der Programmiersprache abhängigen Struktur vorliegen, wird zur Erklärung trotzdem noch die Datei-Struktur verwendet.

## Datennormalisierung

### 6-Spalten-Umsteiger

Sowohl die ICD-10-GM Versionen 2004 und 2.0, als auch die OPS Version 2.1 beinhalteten Umsteiger in folgendem Format:

A00.0;A00.0;A;A;O;UNDEF
1-202;1-202;A;A;;

Um diese an das ICD-10-GM Format von 2024 anzupassen, werden die letzten beiden Spalten entfernt.

### OPS Umsteiger

Für OPS Versionen ab 2010 sehen die Umsteiger-Einträge so aus:

1-100;N;1-100;N;A;A
---------------------

Durch Entfernung der zweiten und vierten Spalte stimmen diese mit den ICD-10-GM Umsteiger Format von 2024 überein.

### OPS 6-Spalten-Umsteiger, altes Format

Die OPS Versionen 2009 bis 2006 hatten ebenfalls sechs Spalten für die Umsteiger, aber in einer anderen Reihenfolge:

1-100;1-100;N;N;A;A
---------------------

Hier müssen also die dritte und vierte Spalte entfernt werden.

### OPS 5-Spalten-Umsteiger

Die Umsteiger von OPS Version 2005 waren in einem ganz eigenen Format geschrieben:

5-062.0;5-062.0;N;A;A
5-062.1;5-062.1;N;A;A
5-062.2;5-062.8;J;E;E
5-062.3;5-062.8;J;B;B

Hier wird dritte Spalte entfernt und außerdem werden die Sonderformen für automatische Überleitbarkeit von *B* und *E* nach *A* umbenannt.

## Imputation

Für Umsteiger der OPS Version 2.0 gibt es nur drei Spalten:

1-208.0;A;1-209.0
1-208.x;;1-209.4

Die zweite Spalte zeigt allein die Überleitbarkeit an. Für die Angleichung an das ICD-10-GM Format von 2024 wird also die zweite Spalte entfernt und gedoppelt angehängt. Aus den beiden Beispielzeilen wird damit:

1-208.0;1-209.0;A;A
1-208.x;1-209.4;;

## Datenbereinigung

**KOMBI-Kode** Aus OPS Versionen 2.1 und 2.0 wird die erste Zeile der Kode-Datei entfernt, welche den KOMBI-Eintrag enthält.

**None statt UNDEF** Für OPS Versionen 2009 bis 2004 wird der Kode-Wert None durch UNDEF ersetzt, sowohl in den Kodes-, als auch in den Umsteiger-Dateien.

**Kreuz-Stern-System** Für die ICD-10-GM Versionen 2.0 und 1.3 werden die Zeichen +, \* und ! aus den Kode-Werten entfernt – sowohl in der Kodes-, als auch der Umsteiger-Datei.

**Punkt-Strich-Notation** Für die ICD-10-GM Versionen 2004, 2.0 und 1.3 wird zuerst die Zeichenfolge .- aus den Kode-Werten in beiden Dateitypen entfernt. Danach wird nochmals - entfernt. Die Reihenfolge ist wichtig, weil in Version 2004 zum Beispiel Kodes A00.- und G82.1- vorkommen.

## Instanzreduktion

### Umsteiger

Für viele Versionen sind über 90% der Umsteiger-Einträge beidseitig automatische Überleitungen in den gleichen Kode. Diese müssen also gar nicht in eine Applikation aufgenommen werden, unter der Annahme, dass nicht vorhandene Umsteiger damit automatischen Überleitungen entsprechen. Dadurch können alle Umsteiger ausgeschlossen werden, bei denen der neue Kode gleich dem alten Kode ist und die automatische Überleitbarkeit in beide Richtungen gegeben ist.

### Nicht-endständige Umsteiger

Die Überleitung von ICD-10-GM Version 2.0 auf 1.3 ist der einzige Fall, in dem die Umsteiger-Datei nicht-endständige Codes enthält. Durch das Entfernen der Sonderzeichen von Punkt-Strich-Notation und Kreuz-Stern-System gibt es außerdem doppelte Einträge in der ersten Spalte, also bei den Codes, die sich auf die Vorgänger-Version beziehen.

Folgende Funktion<sup>1</sup> entfernt die überflüssigen Einträge.

Angenommen die Codes befinden sich in einem Array mit Index von 0 bis (n-1), zum Beispiel für die ersten sechs Zeilen der Umsteiger-Datei aus ICD-10-GM, Überleitung 2.0 nach 1.3:

Index	old (Alter Kode)	new (Neuer Kode)
0	A00	A00.0
1	A00	A00.1
2	A00	A00.9
3	A00.0	A00.0
4	A00.1	A00.1
5	A00.9	A00.9

### removeNonTerminal

Funktionsparameter:

- \$data  
Umsteiger-Einträge

Lokale Variablen:

- \$index  
Der Umsteiger-Eintrag, der aktuell verarbeitet wird.  
Die Funktion läuft vom letztem Eintrag zum ersten.
- \$current  
Der alte Kode des aktuellen Umsteiger-Eintrags.  
Hierbei bedeutet \$data[ \$index ][ 'old' ] Zugriff auf Zeile mit Index = Wert von \$index und Spalte 'Alter Kode'. Zum Beispiel: data[ 4 ][ 'old' ] = A00.1.
- \$prev  
Der alte Kode des zuletzt verarbeiteten Umsteiger-Eintrags.  
Die Variable wird auf \$current gesetzt, wenn der Eintrag *nicht* entfernt wird.

<sup>1</sup>Der Pseudocode ist beschrieben in Struktogrammen nach (Nassi und Shneiderman, 1973). Die später vorkommenden Algorithmen zur Umsteiger-Suche enthalten viele Variablenzuweisungen, die abhängig von der Suchrichtung sind und deren parallele Darstellung in Nassi-Shneidermann-Diagrammen übersichtlicher ist. Sie werden also im Sinne der Einheitlichkeit ebenfalls für den Algorithmus in diesem Kapitel verwendet.

## 2 BfArM-Daten

\$index = count(\$data) - 1	
\$current = \$data[ \$index ][ 'old' ]	
WHILE \$index > 0	
\$prev = \$data[ \$index-1 ][ 'old' ]	
IF \$current contains \$prev AND length(\$current) > length(\$prev)	
Y	N
remove(\$data[ \$index-1 ])	\$current = \$prev
\$index = \$index - 1	
reindex(\$data)	

Die IF-Bedingung bezieht sich auf String-basierte Operationen; also contains: „enthält“ als Sub-String und length: Anzahl der Zeichen im String. remove entfernt eine ganze Zeile, beziehungsweise einen ganzen Umsteiger-Eintrag.

Beispiel-Array nach Durchlaufen des Algorithmus:

Index	old (Alter Kode)	new (Neuer Kode)
0	A00.0	A00.0
1	A00.1	A00.1
2	A00.9	A00.9

### Nicht-endständige Kodes

Für die meisten Anwendungen sind eigentlich nur die endständigen Kodes relevant. Statt diese bei jeder Operation herauszufiltern, können beim einmaligen Einlesen der Daten auch einfach die nicht-endständigen Kodes ausgeschlossen werden. Zu diesem Zweck kann der oben erklärte Algorithmus ebenfalls eingesetzt werden.

### Integration zusätzlicher Informationen

Das nächst Kapitel erklärt, wie ermittelt wird, ob es zu einem Kode einer bestimmten Version Umsteiger in einer älteren oder neueren Version gibt. Um diese zusätzliche Information speichern zu können, werden die Kodes um eine Spalte erweitert.



# 3

## Umsteiger-Suche

Dieses Kapitel behandelt die versionsübergreifende Suche nach Umsteigern. Es geht darum zu bestimmen, ob und wie ein bestimmter Kode in Betrachtung auf alle Versionen des Kodiersystems übergeleitet wurde. Die hier aufgeführten Algorithmen sind unabhängig von einer konkreten Implementation verfasst – als Datenstrukturen werden nur assoziative Arrays, also Maps mit Key⇒Value Paaren, sowie gewöhnliche Arrays mit Index verwendet.

### 3.1 Allgemeine Funktionen

Grundlegend ist davon auszugehen, dass Funktionen zum Lesen der Daten vorliegen – nach erfolgreicher Integration aus dem vorherigen Kapitel.

#### Daten Lesen

Konkret könnten zum Beispiel die Inhalte der Kodes- und Umsteiger-Dateien mit folgenden Parametern ausgelesen werden.

##### **readData**

- `$system`  
Das Kodiersystem, beispielsweise als Konstanten definiert: 'icd10gm' und 'ops'.
- `$version`  
Die Version, beziehungsweise Jahreszahl.
- `$code`  
Ein Kode, nach dem gesucht wird. In Bezug auf die Notation von ICD-10-GM und OPS ist es sinnvoll, automatisch an rechter Stelle einen Wildcard zu implizieren – das heißt einen Platzhalter für beliebige, weitere Zeichen. Dann werden mit einem leeren Wert alle Daten für die anderen Parameter gelesen ohne Einschränkung auf den Kode.

- \$type

Die Art der Daten, die gelesen werden sollen. Möglich ebenfalls als Konstanten:

- 'kodes'

Kodes-Daten: Kode und Titel.

- 'umsteiger'

Umsteiger-Daten: alter Kode und neuer Kode, sowie Information über die automatische Überleitbarkeit in beide Richtungen. Es werden die Umsteiger von der angegebenen auf die nächstältere Version abgefragt. Suche über \$code = neuer Kode.

- 'umsteiger\_join'

Wie Umsteiger-Daten, aber zusätzlich mit Titel für jeweils den alten und neuen Kode. „Join“ bezieht sich auf die entsprechende Datenbankoperation.

- 'umsteiger\_join\_alt'

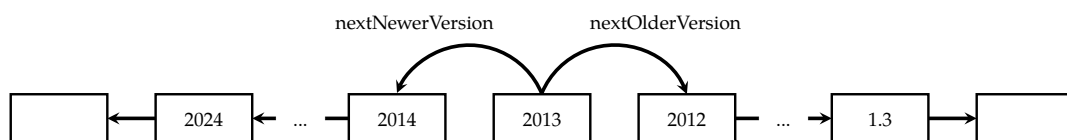
Wie oben, außer dass die Suche über \$code = alter Kode und in die andere Richtung erfolgt. Das heißt es wird die angegebene Version mit der nächstneueren Version verglichen. „Alt“ für „alternate direction“.

## Nächste Version

Außerdem wird eine Funktion benötigt, die ausgehend von System und Version die jeweils nächstältere oder -neuere Version ermittelt. Hierfür kann die in Unterabschnitt 2.1 erwähnte strukturierte Datei dienen, welche alle Versionen und Abweichungen definiert.

Wenn keine neuere oder ältere Version existiert, bleibt der Rückgabewert leer.

### nextNewerVersion / nextOlderVersion



**Abbildung 3.1:** Ermitteln der nächstliegenden Versionen am Beispiel von ICD-10-GM.

## Beispieldaten mit graphischer Darstellung

Um beispielsweise die Codes M21.4 und M21.6 aus der ICD-10-GM Version 2016 mit der Funktion readData abzufragen, sehen die Aufrufe so aus:

```
readData ('icd10gm', '2014', 'M21.4', 'kodes')
```

```
readData ('icd10gm', '2014', 'M21.6', 'kodes')
```

Die Ergebnisse in Tabellenform:

code (Kode)	name (Titel)
M21.4	Plattfuß [Pes planus] (erworben)
M21.6	Sonstige erworbene Deformitäten des Knöchels und des Fußes

Für das restliche Kapitel wird angenommen, dass die Ergebnisse der readData-Funktion in einer zweidimensionalen Datenstruktur gespeichert werden, die einer Tabelle ähnelt – das heißt eine Array-Liste mit einem Eintrag pro Zeile beginnend mit Index = 0, sowie einem assoziativen Array (Map oder alternativ eine Klassenstruktur) mit Key = Spaltenüberschrift. Der Zugriff auf den Titel „Plattfuß“ in deren oberen Tabelle erfolgt dann über `$data[ 0 ][ 'name' ]`.

(Mit dem \$-Zeichen werden Variablennamen hervorgehoben.)

Analog zu den Codes können die Umsteiger wie folgt abgefragt werden:

```
lies_daten ('icd10gm', '2014', 'M21.4', 'umsteiger')
lies_daten ('icd10gm', '2014', 'M21.6', 'umsteiger')
```

Resultierend in:

new (Kode Version 2014)	old (Kode Version 2013)	←Auto→	
M21.4	M21.4	A	A
M21.6	M21.6	A	A

“Auto” steht für die automatische Überleitbarkeit in die jeweilige Richtung, das heißt 2014←2013 und 2014→2013.

Von Version 2014 auf 2013 ändern sich die angegebenen ICD-10-GM Codes nicht. Sie sind auch in beide Richtungen automatisch überleitbar. Wie bereits in der Datenintegration erwähnt, müssen diese Einträge eigentlich gar nicht aufgenommen werden. Sie sind hier nur zur Vereinfachung aufgelistet, weil die Einträge auch so in den BfArM-Dateien enthalten sind. Bei der versionsübergreifenden Suche nach Umsteigern sind allerdings nur die Veränderungen relevant. Anstatt alle nicht relevanten Umsteiger-Einträge abzuspeichern und bei jeder Abfrage auszuschließen, kann alternativ auch angenommen werden, dass bei Nichtvorhandensein eines Umsteigers zwischen zwei Versionen dieser Code einfach gleich bleibt. Die Suche nach Umsteigern geht immer von einem vorhanden Code aus.

*Wichtig:* Für den Rest des Kapitels bedeutet Umsteiger, dass bei der Überleitung eines Kodes von der alten auf die neue Version wirklich eine Veränderung vorliegt.

### 3 Umsteiger-Suche

M21.4 hat also keine Umsteiger. M21.6 hat folgende:

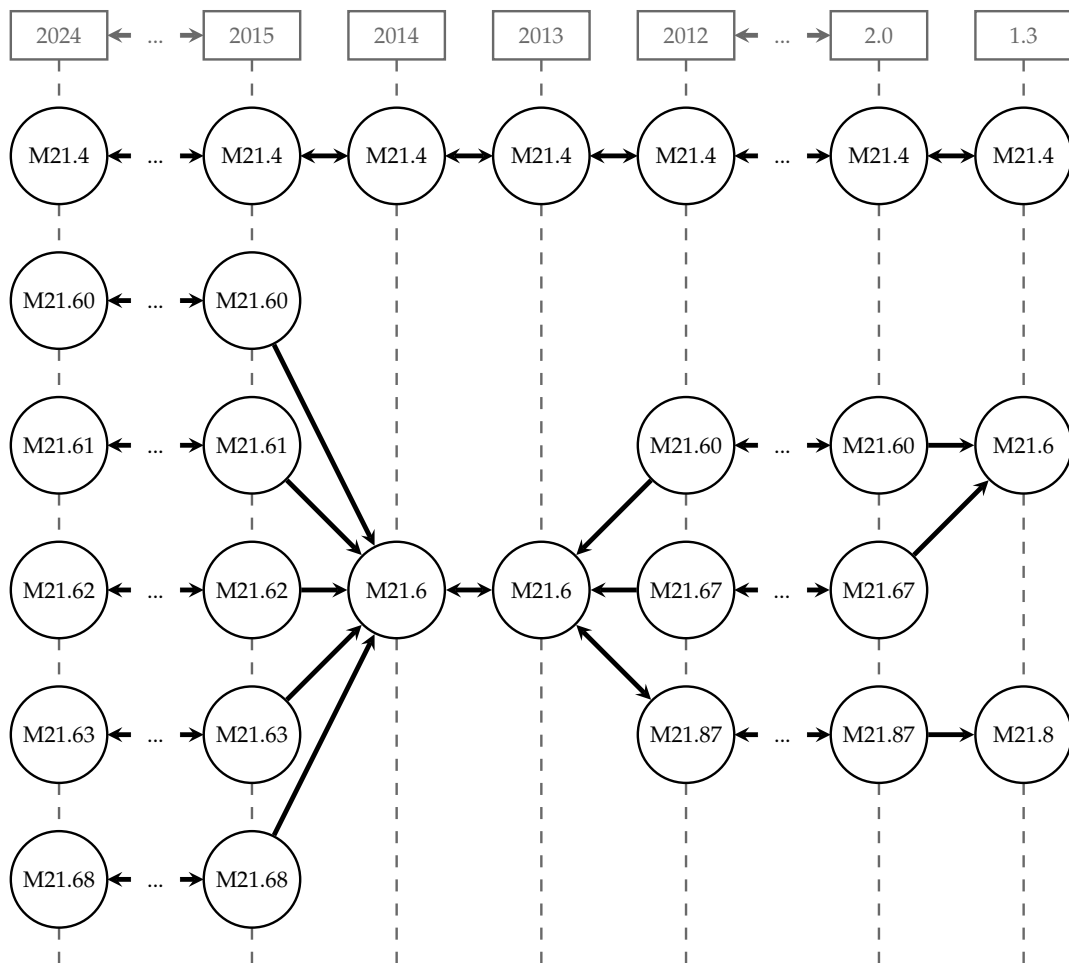
new (Kode Version 2015)	old (Kode Version 2014)	←Auto→	
M21.60	M21.6	A	
M21.61	M21.6	A	
M21.62	M21.6	A	
M21.63	M21.6	A	
M21.68	M21.6	A	
new (Kode Version 2013)	old (Kode Version 2012)	←Auto→	
M21.6	M21.60	A	
M21.6	M21.67	A	
M21.6	M21.87	A	A

Die Suche wird ab Version 2012 mit M21.60, M21.67 und M21.87 fortgesetzt:

new (Kode Version 2.0)	old (Kode Version 1.3)	←Auto→	
M21.60	M21.6	A	
M21.67	M21.6	A	
M21.80	M21.8	A	
M21.81	M21.8	A	
M21.82	M21.8	A	
M21.83	M21.8	A	
M21.84	M21.8	A	
M21.85	M21.8	A	
M21.86	M21.8	A	
M21.87	M21.8	A	
M21.88	M21.8	A	
M21.89	M21.8	A	A

Anmerkung: Die Umsteiger für Kode M21.8, Version 1.3. sind nur zur Vollständigkeit gelistet. Von M21.6, Version 2014 ausgehend ist nur M21.87, Version 2.0 erreichbar.

Die Ergebnisse können als Graph dargestellt werden; siehe nächste Seite. Die Versionen sind horizontal angeordnet und die Kodes pro Version vertikal darunter als Knoten. Ausgangspunkt sind wie erwähnt die Kodes M21.4 und M21.6 der Version 2014. Die Knoten der anderen Versionen entsprechen den Kodes, die über Umsteiger erreichbar sind. Die Pfeilspitzen geben die Richtung der automatischen Überleitbarkeit an.



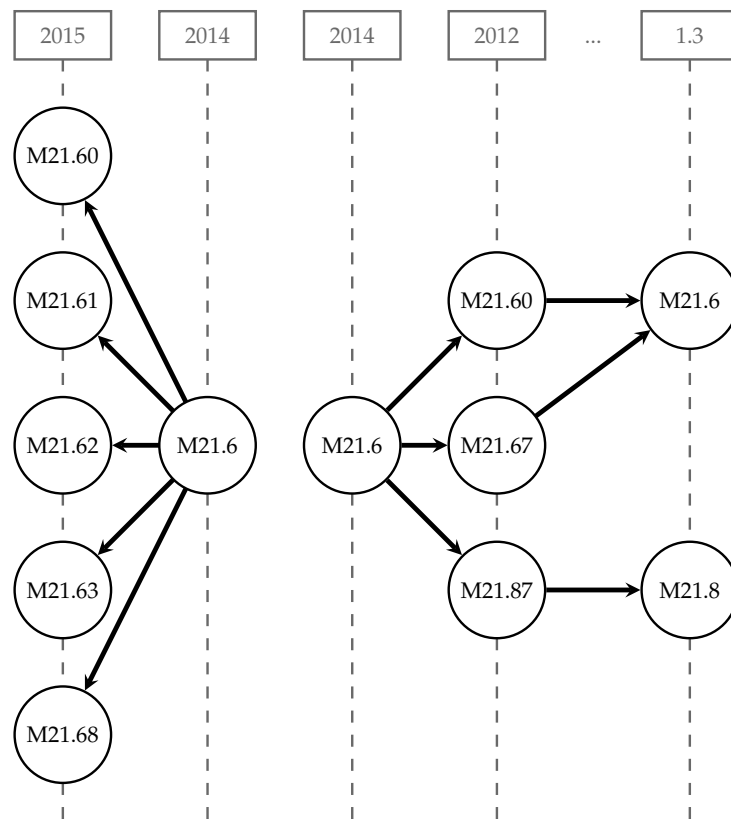
**Abbildung 3.2:** Graphische Repräsentation der Überleitungen ausgehend von den Codes M21.4 und M21.6 der ICD-10-GM Version 2014.

## 3.2 Transitive Hülle

Wenn statt der graphischen Repräsentation oben:

1. Die nicht relevanten Überleitungen ausgeschlossen werden; also wie besprochen nur Umsteiger-Einträge mit Veränderung zählen.
2. Die Kanten im Graph die Suchrichtung anzeigen statt der automatischen Überleitbarkeit.

Dann ergibt sich ein gerichteter Graph – das heißt ein Graph, der nur einfach gerichtete Kanten enthält und in dem jeder Knoten mit mindestens einem anderen Knoten über eine gerichtete Kante –auch Bogen genannt– verbunden ist. Unter diesen Voraussetzungen entspricht die versionsübergreifende Suche nach Umsteigern der Bestimmung der transitiven Hülle in der Graphentheorie.



**Abbildung 3.3:** Von M21.6, Version 2014 über Umsteiger in chronologische und rückwärts chronologische Richtung erreichbare Codes als zwei gerichtete Graphen.

Die transitive Hülle wird in (Gross, Yellen und Zhang, 2013, Seite 172) wie folgt definiert:

**D26:** The *transitive closure*  $R^*$  of a binary relation  $R$  is the relation  $R^*$  defined by  $(x, y) \in R^*$  if and only if there exists a sequence  $x = v_0, v_1, v_2, \dots, v_k = y$  such that  $k \geq 1$  and  $(v_i, v_{i+1}) \in R$ , for  $i = 0, 1, \dots, k - 1$ . Equivalently, the transitive closure  $R^*$  of the relation  $R$  is the smallest transitive relation that contains  $R$ .

**D27:** Let  $G$  be the digraph representing a relation  $R$ . Then the digraph  $G^*$  representing the transitive closure  $R^*$  of  $R$  is called the *transitive closure of the digraph*  $G$ . Thus, an arc  $(x, y)$ ,  $x \neq y$ , is in the transitive closure  $G^*$  if and only if there is a directed  $x$ - $y$  path in  $G$ . Similarly, there is a self-loop in digraph  $D^*$  at vertex  $x$  if and only if there is a directed cycle in digraph  $G$  that contains  $x$ .

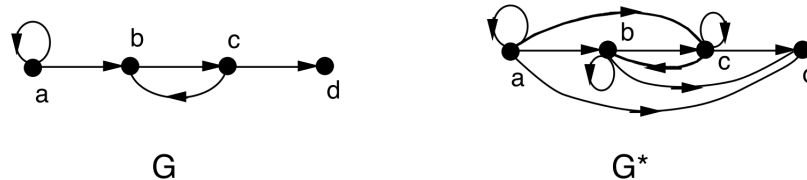
Oder kurz zusammengefasst: die transitive Hülle eines gerichteten Graphens  $G$  ist wiederum ein gerichteter Graph  $G^*$ , der von jedem Knoten einen Bogen zu allen von diesem Knoten aus in  $G$  erreichbaren Nachbarn besitzt.

In (Gross, Yellen und Zhang, 2013, Seite 172) ist hierzu folgendes Beispiel dargestellt:

**E7:** Suppose a relation  $R$  on the set  $S = \{a, b, c, d\}$  is given by

$$\{(a, a), (a, b), (b, c), (c, b), (c, d)\}$$

Then the digraph  $G$  representing the relation  $R$  and the transitive closure  $G^*$  are as shown in Figure 3.1.7.



Es gibt mehrere Verfahren, um die transitive Hülle zu bestimmen. Es werden nun zwei Algorithmen vorgestellt, die anhand ihrer primären Suchrichtung benannt sind.

Wie in den Graphen dargestellt, läuft die *horizontale* Suche primär über die Versionen – mit jeweils einzelnen Codes. Die *vertikale* Suche läuft primär über die Codes – das heißt es werden zuerst alle Codes einer Version gelesen.

#### Horizontale Suche

Ein intuitives Verfahren die transitive Hülle zu bestimmen wird in (Jakobsson, 1991, Seite 200) so beschrieben:

##### 2.2 Search

A second approach for computing the transitive closure is to search the graph  $n$  times, each time starting from a different node, thereby determining what can be reached from that node. In this approach, we completely disregard any parts of the  $TC$ -relation that have already been computed for other starting nodes. Instead, we search the entire part of the graph that is reachable from the starting node by following every edge we can get to.

Ähnlich wie im vorherigen Abschnitt für den Code M21.6 dargestellt, werden also einfach ausgehend von einem Knoten so viele Suchen gestartet bis keine benachbarten Knoten mehr gefunden werden – oder im Anwendungsfall der Umsteiger bis alle Versionen verarbeitet wurden.

Vorher erzielte Ergebnisse für andere Knoten, beziehungsweise Codes, werden ignoriert. Das hieße im Beispiel oben, dass eine rückwärts chronologische Suche nach Umsteigern von M21.60 und M21.61 der ICD-10-GM Version ab dem Vorgänger M21.6 der Version 2014 zweimal exakt gleich ablaufen würde.

Der Pseudocode für diese Vorgehensweise:

#### **searchHorizontal**

Funktionsparameter:

- \$system, \$version, \$code  
Wie Funktion readData, siehe Unterabschnitt 3.1.

Lokale Variable:

- \$data      Rückgabewert, initial: leer.  
Eine assoziative Datenstruktur mit Key fwd  $\Rightarrow$  Umsteiger, die chronologisch vorwärts von dem Suchkode erreichbar sind und rev  $\Rightarrow$  chronologisch rückwärts.

<code>\$data[ 'fwd' ] = searchHorizontalRecursion (\$system, \$version, \$code, true)</code>
<code>\$data[ 'rev' ] = searchHorizontalRecursion (\$system, \$version, \$code, false)</code>
<code>RETURN \$data</code>

Ausgehend von einem gegebenen Kodiersystem, Version, Kode wird chronologisch vorwärts und rückwärts eine rekursive Suche über alle Versionen gestartet.

#### **searchHorizontalRecursion**

Funktionsparameter:

- \$system, \$version, \$code  
Wie searchHorizontal.
- \$chronological  
Bestimmt die Suchrichtung. TRUE: chronologisch vorwärts und FALSE: rückwärts.

Lokale Variablen:

- \$data      Rückgabewert, initial: leer.  
Eine mehrdimensionale, assoziative Datenstruktur, die die Ergebnisse der rekursiven Suche enthält. Gefundene Umsteiger werden als Liste mit dem Key umsteiger gespeichert; zusätzlich zu den Listen wird auch die Versionen vorher und nachher notiert. Jeder einzelne Umsteiger-Eintrag enthält wiederum die Informationen zur automatischen Überleitbarkeit, sowie die Titel der Kodes vorher und nachher. Falls ein Umsteiger weitere Umsteiger hat, werden diese unter einem Key recursion abgelegt und die Suche fortgesetzt. Das heißt unter recursion ist das gleiche Datenkonstrukt nochmal enthalten. Ein Beispielergebnis für M21.6 befindet sich im Anhang A.2.
- \$otherVersion  
Die Version, zu der die Umsteiger ermittelt werden; \$version  $\rightarrow$  \$otherVersion.
- \$otherCode  
Die Spalte im Umsteiger-Eintrag, zu der der aktuelle Kode übergeleitet wird.



### 3 Umsteiger-Suche

- \$readType  
Auf welche Art die Daten gelesen werden; siehe readData.
- \$umsteiger, \$entry  
Liste der Umsteiger, Schleifenvariable: Umsteiger-Eintrag.
- \$recursion  
Ergebnis des rekursiven Funktionsaufrufs bei gefundenen Umsteigern.

IF \$chronological	
Y	N
\$otherVersion = nextNewerVersion(\$system, \$version)	\$otherVersion = nextOlderVersion(\$system, \$version)
\$readType = 'umsteiger_join_alt'	\$readType = 'umsteiger_join'
\$otherCode = 'new'	\$otherCode = 'old'

Lokale Variablen werden anhand der Suchrichtung unterschiedlich belegt.

IF empty(\$version) OR empty(\$otherVersion)	
Y	
RETURN \$data	∅

Falls es in der Suchrichtung keine weitere Version mehr gibt, endet die Rekursion.

\$umsteiger = readData(\$system, \$version, \$code, \$readType)	
IF empty(\$umsteiger)	
Y	
RETURN searchHorizontalRecursion(\$type, \$otherVersion, \$code, \$chronological)	∅

Die Umsteiger zwischen den Versionen im aktuellen Rekursionsschritt werden ermittelt. Falls es keine gibt –also keine, die eine Veränderung ausdrücken – dann wird die Rekursion mit der nächsten Version fortgesetzt.

FOREACH \$entry IN \$umsteiger		
IF \$entry[ \$otherCode ] IS NOT 'UNDEF'		
Y		
\$recursion = searchHorizontalRecursion (\$system, \$version, \$entry[ \$otherCode ], \$chronological)		∅
IF NOT empty(\$recursion)		
Y		
\$entry[ 'recursion' ] = \$recursion	∅	
\$data[ 'umsteiger' ][ ] = \$entry		

Für jeden Umsteiger wird die Rekursion mit dem veränderten Code fortgesetzt. Allerdings nur falls dieser nicht UNDEF ist, weil es sich dann um einen entfer-

ten oder neu hinzugefügten Kode handelt. Also wenn wie im Beispiel der M21.6 der ICD-10-GM Version 2013 die Umsteiger [ M21.60, M21.67, M21.87 ] in der Version 2012 hat, dann wird die Suche in die rückwärts chronologische Richtung mit diesen drei Kodes fortgesetzt statt mit M21.6. Die Ergebnisse dieser Verzweigung, sofern vorhanden, werden abgespeichert mit dem Key `recursion` zusätzlich zu jedem Umsteiger-Eintrag. Die Liste aller Umsteiger-Einträge, inklusive der UNDEF-Umsteiger, wird in das Ergebnis mit dem Key `umsteiger` aufgenommen.

<code>\$data[ 'version' ] = \$version</code>
<code>\$data[ 'other' ] = \$otherVersion</code>
<code>RETURN \$data</code>

Falls es Umsteiger für einen Kode gibt, werden zusätzlich die Versionen vorher und nachher gespeichert. Dann endet auch hier die Rekursion.

Anhang A.2 enthält das Ergebnis im JSON-Format für den Beispielaufwurf der horizontale Suche für M21.6, Version 2014 entsprechend der Abbildung 3.3.

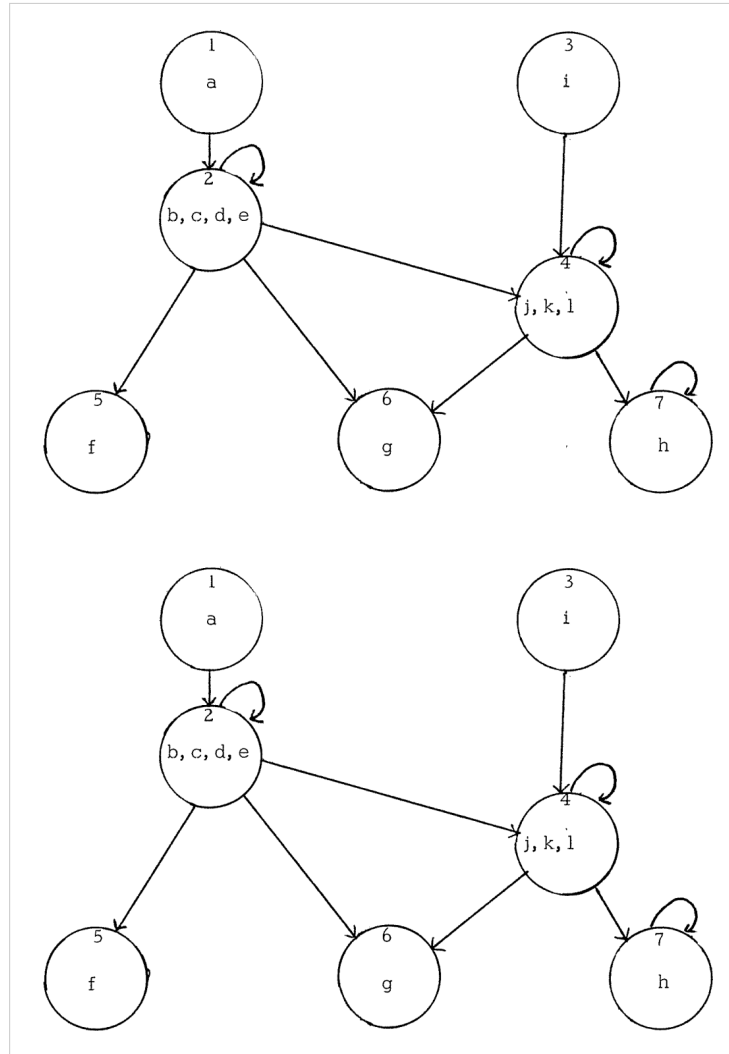
### Vertikale Suche

Wie bereits erwähnt beginnt der Algorithmus der horizontalen Suche für jeden Kode ohne Vorkenntnisse über schon getätigte Aufrufe. Wenn die Umsteiger für alle Kodes einer Version –beziehungsweise eines ganzen Kodiersystems– gefunden werden sollen, ist diese Vorgehensweise nicht effizient.

Besonders für gerichtete Graphen ist allerdings Purdoms Algorithmus gut geeignet, um die transitive Hülle zu bestimmen, wie in (Dar, 1993, Seite 77) beschrieben. Folgende kurze Zusammenfassung des Algorithmus basiert auf dieser Beschreibung, sowie auf (Purdom, 1970) selbst:

1. Ausgehend von einem Graphen  $G$ : Bestimme die stark zusammenhängenden Komponenten (SCC) in  $G$  und vereinige diese in jeweils einen einzelnen Knoten. Das Ergebnis ist ein zyklensfreier, verdichteter Graph  $G_c$ .
2. Sortiere  $G_c$  topologisch.
3. Ermittle die transitive Hülle von  $G_c$  über dessen Adjazenzlisten in rückwärts topologischer Reihenfolge.
4. Bestimme die transitive Hülle des ursprünglichen Graphens  $G$  über die Nachbarschaftslisten der stark zusammenhängenden Komponenten in  $G_c$ : ein Knoten  $y$  ist ein erreichbarer Nachbar von Knoten  $x$ , falls  $SCC(y) = SCC(x)$  oder falls  $SCC(y)$  ein erreichbarer Nachbar von  $SCC(x)$  ist.

Die Vereinigung der stark zusammenhängenden Komponenten aus Schritt eins wird in (Purdom, 1970) wie folgt dargestellt:



**Abbildung 3.4:** Beispielgraph G und Vereinigung der stark zusammenhängenden Komponenten aus (Purdom, 1970, Seite 78).

Die hier vorgestellte vertikale Suche nach Umsteigern entspricht nicht ganz Purdoms Algorithmus. Sie weicht in mehreren Punkten ab:

- Die topologische Sortierung aus Schritt zwei in der Anwendung der Suche nach Umsteigern ist nicht notwendig, weil die Versionen immer in einer bestimmte Richtung verglichen werden und die Codes sich nur zwischen zwei Versionen ändern können. Die Daten sind also schon implizit topologisch sortiert.
- Ebenfalls ist nicht notwendig, die komplette transitiven Hülle zu ermitteln, sondern nur wie Codes der Startversion in die Zielversion übergeleitet werden. Die Zwischenergebnisse werden trotzdem gespeichert.
- Auch bei der Suche nach Umsteigern nur mit Veränderungen ist es möglich, anders als bei den Knoten in einem Graphen, dass Codes mehrmals vorkommen. Zum Beispiel gibt es G83.3 in ICD-10-GM nicht zwischen Version 2017 und 2005, aber schon davor und danach.

Trotzdem basiert die vertikale Suche wesentlich auf zwei Besonderheiten von Purdoms Algorithmus, die in (Dar, 1993, Seite 76f) hervorgehoben werden:

### 5.3.1 The Purdom Algorithm

Purdom, in [Pur70], made two key observations:

1. During the computation of transitive closure of a directed acyclic graph, if node  $s \prec_t t$ , then additions to the successor list of node  $s$  cannot affect the successor list of node  $t$ . One should therefore compute the successor list of  $t$  first and then that of  $s$ . By processing nodes in reverse topological order, one need add to a node only the successor lists of its immediate successors since the latter would already have been fully expanded. We call this idea the *immediate successor optimization* [AgJ90].
2. All nodes within a strongly connected component (SCC) in a graph have identical reachability properties, and the condensation graph obtained by collapsing all the nodes in each strongly connected component into a single node is acyclic.

Konkret bedeuten diese Besonderheiten für die Umsteiger-Suche:

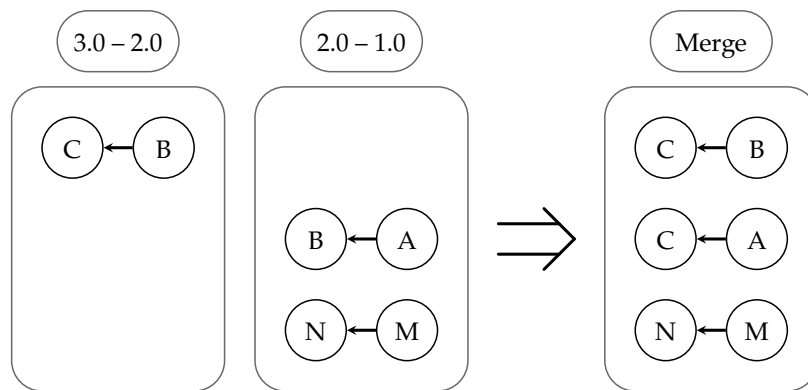
1. Topologisch rückwärts gerichtete Vorgehensweise:  
So ist zum Beispiel bei der Suche nach den Umsteigern eines ICD-GM-10 Kodes der Version 1.3 die Reihenfolge, in der die anderen Versionen abgearbeitet werden: 2024, 2023, 2022 und so weiter. Oder anders ausgedrückt: Wenn die Überleitungen in chronologischer Reihenfolge bestimmt werden sollen, dann läuft die vertikale Suche umgekehrt, also chronologisch rückwärts.
2. Vereinigung der stark zusammenhängenden Komponenten:  
Nach jedem Schritt über eine Version, wird geprüft, ob ein neu gefundener Umsteiger eine Überleitung in einen Code enthält, der bereits in den davor abgearbeiteten Versionen als Umsteiger vorkommt. Falls ja, dann werden diese Umsteiger vereinigt, das heißt die Zwischenschritte werden zu einem Schritt zusammengefügt.

Abbildung 3.5 auf der nächsten Seite zeigt ein einfaches Beispiel für die Vereinigung:

Angenommen ein abstraktes System hat Umsteiger von der Version 1.0 auf 2.0:  $A \rightarrow B$  und  $M \rightarrow N$ , sowie von Version 2.0 auf 3.0:  $B \rightarrow C$ . Wenn nun in chronologischer Reihenfolge  $1.0 \rightarrow 2.0 \rightarrow 3.0$  Umsteiger ermittelt werden sollen, erfolgt die Suche sowie Vereinigung gefundener Umsteiger in umgekehrter Reihenfolge. Das heißt wenn im Schritt  $3.0 \leftarrow 2.0$  der Umsteiger  $C \leftarrow B$  gefunden wird, dann wird danach bei der Vereinigung mit Version 1.0 der Umsteiger  $B \leftarrow A$  in  $C \leftarrow A$  umgewandelt.  $A$  in Version 1.0 entspricht  $C$  in Version 3.0.

Falls es Verzweigungen gibt, also es mehrere Umsteiger für einen Code gibt, dann werden entsprechend Listen vereinigt. Ein reales Beispiel dazu folgt später.

### 3 Umsteiger-Suche



**Abbildung 3.5:** Graphische Veranschaulichung der zwei Besonderheiten von Purdoms Algorithmus, die in der vertikalen Suche zur Anwendung kommen.

Der gesamte Algorithmus für die vertikale Suche als Pseudocode:

#### searchVertical

Funktionsparameter:

- \$system  
Das Kodiersystem.
- \$targetVersion  
Die Zielversion, auf die von allen anderen Version die Umsteiger über alle Codes ermittelt werden sollen.
- \$function  
Eine optionale, anonyme Funktion.

Lokale Variable:

- \$data      Rückgabewert, initial: leer.  
Eine zweidimensionale, assoziative Datenstruktur – wird nur befüllt wenn \$function nicht gesetzt ist. Nach Durchlauf des Algorithmus enthält die erste Dimension als Keys alle Versionen außer der Zielversion. Die zweite Dimension hat als Key jeweils einen Code und als Value die Liste der zugehörigen Umsteiger nach erfolgter Vereinigung. Codes ohne Umsteiger werden nicht aufgenommen.

\$data += searchVerticalSubroutine(\$system, \$targetVersion, false, \$function)
\$data += searchVerticalSubroutine(\$system, \$targetVersion, true, \$function)
RETURN \$data

Basierend auf der Zielversionen wird in beide Richtungen die vertikale Suche als Unterfunktion gestartet.

$\$data+=$  bedeutet, dass die assoziativen Datenstrukturen zusammengefasst werden. Wenn ein Key auf beiden Seiten der Addition existiert, wird der von der linken Seite für die Summe übernommen. Das kann allerdings hier bei den Versionen als Keys nicht passieren, weil jede Version nur einmal bearbeitet wird.

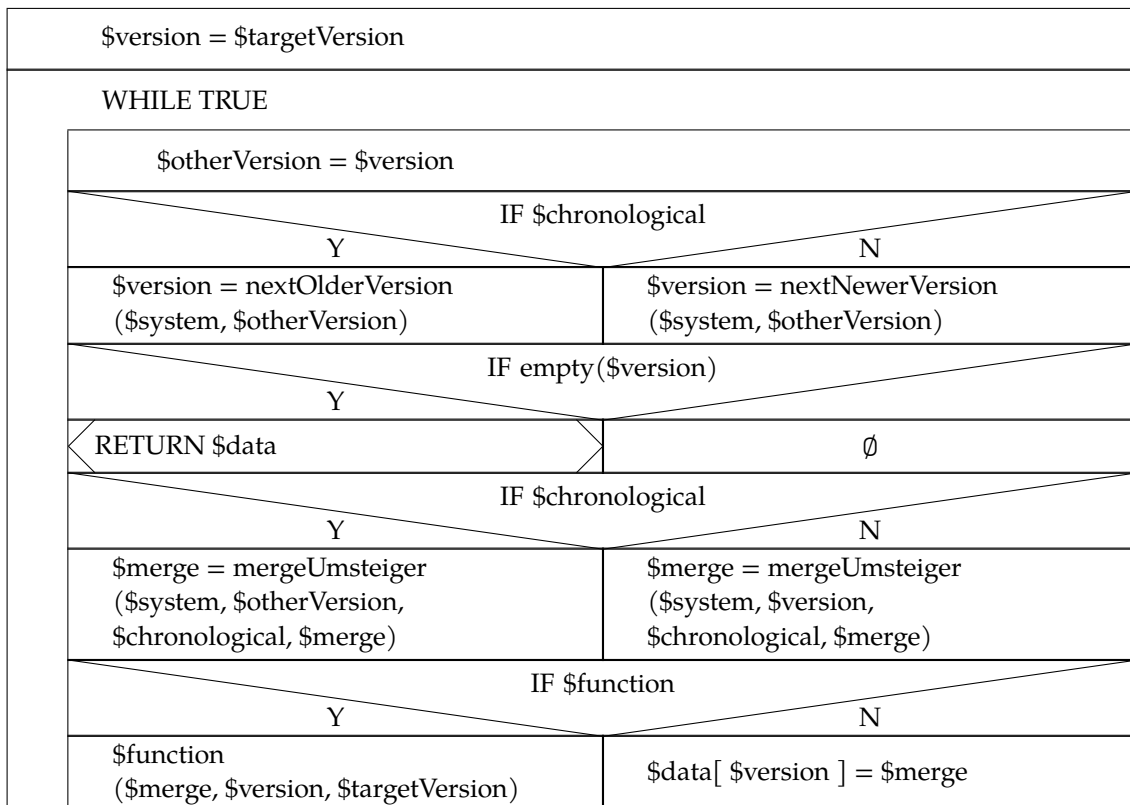
#### searchVerticalSubroutine

Funktionsparameter:

- $\$system, \$targetVersion, \$function$   
Wie searchVertical.
- $\$chronological$   
Bestimmt die Richtung, in der die Versionen abgearbeitet werden. TRUE: chronologisch vorwärts und FALSE: rückwärts. Die Reihenfolge ist wie erwähnt topologisch umgekehrt, das heißt ausgehend von der Zielversion.

Lokale Variablen:

- $\$data$             Rückgabewert, initial: leer.  
Wie searchVertical.
- $\$merge$   
Key-Value-Paare mit Kode  $\Rightarrow$  [übergeleitete Kode(s)] nach Vereinigung.
- $\$version, \$otherVersion$   
Die aktuell zu verarbeitende Version und die nächste Version.



### 3 Umsteiger-Suche

Die Schleife wird solange durchlaufen bis es keine weitere Version in der vorgegebenen Suchrichtung gibt. Pro Durchlauf wird die Variable `$merge` erweitert wie in Abbildung 3.5 dargestellt; sie enthält die vereinigten Umsteiger aller bisher verarbeiteten Versionen. Abhängig davon, ob die anonyme Funktion `$function` definiert ist, wird diese entweder ausgeführt oder die Variable `$data` für die Rückgabe befüllt.

#### mergeUmsteiger

Funktionsparameter:

- `$system, $version, $chronological, $merge`  
Wie `searchVerticalSubroutine`.

Lokale Variablen:

- `$umsteiger`  
Liste aller Umsteiger der Version; initial leer.
- `$item`  
Schleifenvariable der `readData`-Funktion.
- `$current, $other, $currentCode, $otherCode`  
Abhängig von der Richtung werden pro Umsteiger-Eintrag entweder der alte oder neue Code genommen, um die Umsteiger auf die nächste Version zu ermitteln. Im Gegensatz zur horizontalen Suche werden nur die Codes abgefragt – also ohne zusätzliche Informationen. Deswegen ist der `$readType` der `readData`-Funktion hier einfach `'umsteiger'`. Da alle Umsteiger aller Versionen gesammelt werden, könnte die `$merge`-Variable sonst relativ groß werden.

IF \$chronological	
Y	N
<code>\$current = 'old'</code>	<code>\$current = 'new'</code>
<code>\$other = 'new'</code>	<code>\$other = 'old'</code>
FOREACH \$item IN readData(\$system, \$version, "", 'umsteiger')	
<code>\$currentCode = \$item[ \$current ]</code>	
<code>\$otherCode = \$item[ \$other ]</code>	
IF exists(\$merge[ \$otherCode ]) AND NOT ( \$currentCode IS 'UNDEF' OR \$otherCode IS 'UNDEF' )	
Y	N
<code>\$umsteiger[ \$currentCode ] = concatenate(\$merge[ \$otherCode ], \$umsteiger[ \$currentCode ])</code>	<code>\$umsteiger[ \$currentCode ][ ] = \$otherCode</code>
RETURN \$umsteiger + \$merge	

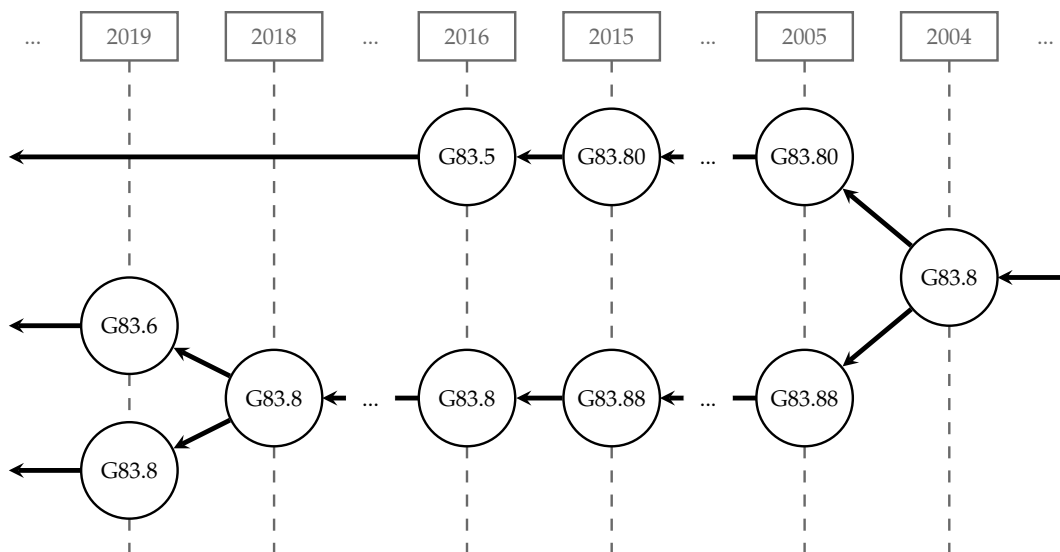
Es werden Umsteiger der aktuellen Versionen gelesen; wie erwähnt nur Überleitungen mit Veränderungen. Wenn der Zielcode einer Änderung bereits Umsteiger in \$merge enthält, der vereinigten Liste vorher verarbeiteter Versionen, dann werden diese genommen und mit den Umsteigern der aktuellen Version vereinigt. concatenate heißt, dass zwei Listen einfach zusammengefügt werden, ohne dass Indizes überschrieben werden. Ansonsten –oder falls Umsteiger UNDEF enthalten, also entfernt oder neu hinzugefügt werden– werden die Umsteiger der aktuellen Version übernommen wie sie sind.

Rückgabewert ist ein assoziatives Array der aktuellen Umsteiger als Key-Value-Paare mit  $\text{Code} \Rightarrow [\text{übergeleitete Code(s)}] + \$merge$ , wobei + hier bedeutet, dass die Einträge von der linken Seite der Summe genommen werden, falls dieselben Keys auf beiden Seiten vorhanden sind.

#### Konkretes Beispiel

Angenommen es soll ermittelt werden, wie der ICD-10-GM Code G83.8 der Version 1.3. in die neueren Versionen übergeleitet wurde.

Der folgende Graph zeigt nur die hierfür relevanten Codes der anderen Versionen. Bei der vertikalen Suche werden immer alle Codes pro Version abgefragt.



**Abbildung 3.6:** Graphisches Beispiel für die vertikale Suche nach Umsteigern ausgehend von ICD-10-GM Code G83.8, Version 1.3 in chronologischer Richtung.

Die Suche erfolgt topologisch rückwärts, das heißt wenn die Suchrichtung chronologisch ist wie im aktuellen Beispiel, dann ist die Bearbeitungsreihenfolge chronologisch rückwärts.

Die Datenstruktur, welche die vereinigten Umsteiger enthält, wird also schrittweise aufgebaut von der neuesten Version ausgehend:



'2018'  $\Rightarrow$  [ 'G83.8'  $\Rightarrow$  [ 'G83.6', 'G83.8' ] ]

'2015'  $\Rightarrow$  [ 'G83.8'  $\Rightarrow$  [ 'G83.6', 'G83.8' ] ],  
              'G83.80'  $\Rightarrow$  [ 'G83.5' ],  
              'G83.88'  $\Rightarrow$  [ 'G83.6', 'G83.8' ] ]

Der erste Eintrag wird übernommen, der zweite kommt neu hinzu und für den dritten Eintrag wird G83.88  $\Rightarrow$  G83.8 entsprechend umgewandelt, weil es Umsteiger für G83.8 schon gibt.

'2004'  $\Rightarrow$  [ 'G83.8'  $\Rightarrow$  [ 'G83.5', 'G83.6', 'G83.8' ] ],  
              'G83.80'  $\Rightarrow$  [ 'G83.5' ],  
              'G83.88'  $\Rightarrow$  [ 'G83.6', 'G83.8' ] ]

Ähnlich wie oben wird hier G83.8  $\Rightarrow$  [G83.80, G83.88] umgewandelt und dann aber der vorhandene Eintrag für G83.8 ersetzt. Das Übernehmen der unveränderten Einträge wird noch bis zur Version 1.3 fortgesetzt, so dass auf der fertigen Datenstruktur gilt:

`$data['1.3']['G83.8'] = [ 'G83.5', 'G83.6', 'G83.8' ]`

### 3.3 Vergleich der Suchalgorithmen

Wie bereits erwähnt erfolgt die horizontale Suche pro Kode über alle Versionen ohne Vorkenntnisse vorheriger Suchergebnisse. Im Gegensatz dazu verarbeitet die vertikale Suche pro Version alle Codes nur einmal, ist also allein deshalb schon effizienter, wenn Umsteiger für mehr als einen Kode ermittelt werden sollen.

Wenn die Daten zusätzlich aus einer Datenbank gelesen werden, ist die Anzahl an Aufrufen der readData-Funktion wesentlich entscheidender als die Menge an Ergebnissen pro Aufruf. Denn das Verbinden mit der Datenbank und Auslösen eines Suchbefehls dauert eine gewisse Zeit, was sich bei sehr vielen Aufrufen schnell aufaddiert. Dementsprechend ist es wesentlich schneller, alle Umsteiger einer Version auf einmal zu lesen, statt einzeln.

Andererseits würde wiederum das Abfragen zusätzlicher Informationen, wie beispielsweise der Titel der Codes eines Umsteigers vorher und nachher, bei der vertikalen Suche schnell die Größe der Daten im Arbeitsspeicher ansteigen lassen, weil pro Aufruf immer alle Umsteiger pro Version zwischengespeichert werden.

Daher werden die beiden Algorithmen für unterschiedliche Zwecke eingesetzt.

Horizontale Suche:

- Darstellung der Umsteiger eines einzelnen Kodes mit Zusatzinformationen

Vertikale Suche:

- Schreiben einer ConceptMap
- Für alle Codes einer Version bestimmen, ob diese Umsteiger haben (ja/nein)

### 3.4 Schreiben der FHIR ConceptMap

Dieser Abschnitt bezieht sich auf die in Kapitel 1.1 beschriebene ConceptMap-Struktur.

Die anonyme Funktion, die der vertikalen Suche übergeben werden kann, dient zum Schreiben der ConceptMap. Jeder Aufruf der Funktion innerhalb des Algorithmus verfügt über die Daten, die zum Schreiben einer *group* an Umsteigern mit Start- und Zielversion benötigt werden. Ein Durchlauf der vertikalen Suche kann dann mehrere Gruppen generieren, die ein Mapping auf eine Zielversionen ausgehend von allen anderen Versionen des Kodiersystems ermöglichen. Ebenfalls kann die Kombination mehrerer vertikalen Suche für jede Version als Zielversion aufgerufen werden, um so eine ConceptMap von allen Versionen auf alle Versionen zu schreiben.

Eine alle Versionen umfassende ConceptMap wird jedoch sehr groß. Zum Beispiel für die ICD-10-GM mit allen Codes aus allen Versionen im XML-Format ist die Datei circa 1 GB groß. Wie in (Braaksma, 2014) beschrieben macht es Sinn solche Dateien zu streamen, denn andernfalls müsste die komplette Datei im Arbeitsspeicher des Servers geschrieben werden. Die Vorgehensweise über die anonymen Funktionen ermöglicht genau das Streaming, weil nach jedem Aufruf der Ausgabepuffer des Servers geleert werden kann. Ein weiterer Vorteil ist, dass selbst wenn ein User sehr große ConceptMaps generieren lässt, der Download-Fortschritt über das Streaming angezeigt wird.

Konkret enthält jede Iteration der anonymen Funktion die Informationen über Zielversion, aktuelle verarbeitete Version, sowie die vereinigten Umsteiger über alle vorher verarbeiteten Versionen. Zusätzlich muss allerdings noch bekannt sein, welche Codes insgesamt in der Startversion vorhanden sind; zum einen, um Codes ohne Umsteiger abzudecken und zum anderen, um keine Umsteiger aus vorher verarbeiteten Version aufzunehmen, deren Ursprungskode in der aktuellen Version nicht existiert. Hierfür ist also ein weiterer `readData`-Aufruf notwendig. Aus dem Vergleich dieser Codes mit den Einträgen in der Merge-Datenstruktur können die Relationen für das Mapping bestimmt werden:

Kode in Merge-Daten enthalten?	Relation (R4)
Nein	<i>equivalent</i>
Ja, übergeleitet auf UNDEF	<i>unmatched</i>
Ja, übergeleitet auf einen Kode	<i>related-to</i>
Ja, überleitetet auf mehrere Kodes	<i>wider</i>

Ablauf des Schreibens einer ConceptMap mit einer einzelnen Zielversion:

1. Anfang der ConceptMap schreiben: *ConceptMap, url, id*.
2. Vertikale Suche mit anonymer Funktion starten; jeder Aufruf generiert eine *group*. Für jeden Kode wird ein *element*→*code*, *element*→*target*→*code* Paar geschrieben – oder bei Relation *wider* entsprechend mehrere *target*→*code* Einträge.
3. Ende der ConceptMap schreiben.

Eine Beispiel-ConceptMap im XML-Format befindet sich im Appendix A.3.

# 4

## Web-Applikation

### 4.1 Überblick

asada bfarmer asd

Version Control: thomas2019pragmatic (data1)

- Daten Integration per Kommandozeilenbefehl
- Darstellung der Daten
- Generieren von ConceptMaps
- Beispiel-Anwendung: Virtuelle Patienten

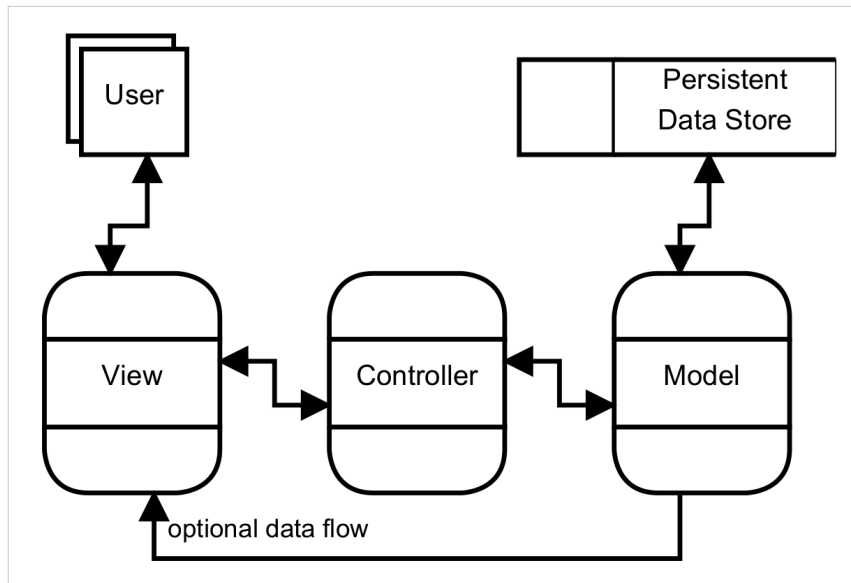
Symfony – Warum? (Potencier, 2022)

Composer (Adermann, Boggiano u. a., o. D.)

PHP Verwendung (W3Techs, o. D.)

Programming Languages Harmful (Janssens, Schultz und Zaytsev, 2017)

Model-View-Controller (Voorhees, 2020, Seite 176f)



**Abbildung 4.1:** (Voorhees, 2020, Seite 177)

Abstraction and Decoupling – Orthogonality: thomas2019pragmatic

Strukturierte XML-Datei

Symfony (Potencier, 2022)

Dependency Injection (Seemann und Deursen, 2019)

### Symfony Komponenten

Console (Symfony, o. D.[a], Console Commands)

Http Foundation: Request, Response, StreamedResponse (Symfony, o. D.[b], The HttpFoundation Component)

HttpClientInterface – Download (Symfony, o. D.[c], HTTP Client)

File System (Symfony, o. D.[d], The Filesystem Component)

UUID (Symfony, o. D.[e], The UID Component)

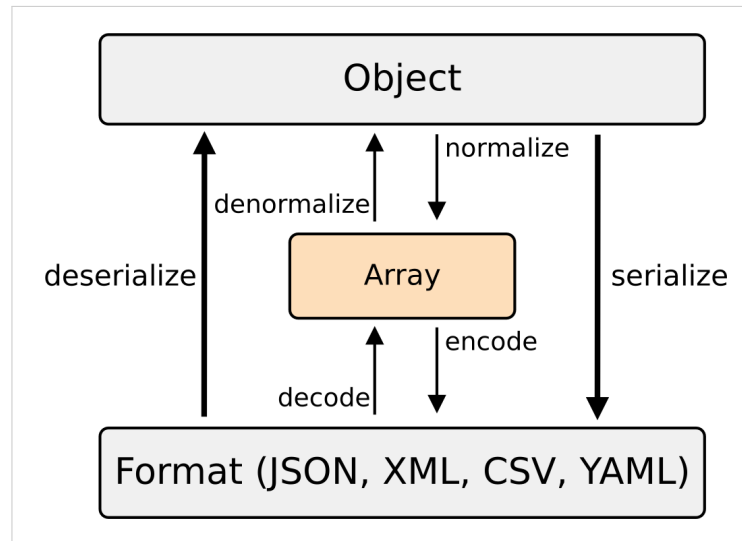
Controller – Routing (Symfony, o. D.[f], Controller)

Serializer: XML, CSV, JSON, SQL-Insert (Symfony, o. D.[g], The Serializer Component)

Database (Symfony, o. D.[h], Databases and the Doctrine ORM)

Templates (Symfony, o. D.[i], Creating and Using Templates)

Twig (Twig, o. D.)



## 4.2 Aufbau des bfarmer -Projekts

### Views

Templates für HTML-Code.

- base  
Grundvorlage für alle Seiten, also zum Beispiel <head> mit Imports für Stylesheets, Javascript und <body>.
- index  
b
- codes  
c
- conceptmap  
d
- modal  
e
- patients  
f
- search\_code  
g
- umsteiger  
h
- umsteiger\_icons  
i

- umsteiger\_search  
j
- umsteiger\_search\_recursion  
k
- umsteiger\_search\_result  
l

### Utility

- Constants  
Enthält globale Konstanten und statische Funktionen, die keine Instanziierung einer Klasse benötigen, um zum Beispiel eindeutige Namen für die SQL-Tabellen zu generieren.
- SqlInsertEncoder  
b
- TwigExtension  
Erweiterung der Twig-Template Engine, um externe Code-Links und

### Repositories

- BfarmRepository  
a
- ConfigRepository  
b
- DatabaseRepository  
c
- PatientsRepository  
d

### Controllers

- AdminController  
a
- CodesController  
b
- ConceptMapController  
c
- DimdiController  
d

- IndexController  
e
- PatientsController  
f
- UmsteigerController  
g
- UmsteigerSearchController  
h

### Services

- ClientService  
a
- ConceptMapService  
b
- DataService  
c
- PatientsService  
d
- SetupService  
e
- UmsteigerService  
f

### Commands

- BfarmerCommand  
a
- TestCommand  
b

## 4.3 Backend

### Setup-XML

### Konsole

### Datenbank / MySQL / Doctrine

Database must have collation utf8mb4\_unicode\_ci

Schreiben der ConceptMap

JSON und XML. Für eine nähere Erklärung dieser Datei-Formate siehe (Bonnefoy u. a., 2024, JSON, Seite 133f) und (Bonnefoy u. a., 2024, XML, Seite 134ff).

### 4.4 Frontend

HTML (WHATWG, o. D.), CSS (W3C, o. D.), JavaScript (Ecma International, o. D.)

#### Webpack / Node.js

Webpack (Koppers u. a., o. D.), NPM (Schlueter, o. D.), Node.js (Dahl, o. D.)

webpack is a static module bundler for modern JavaScript applications

#### Bootstrap

Bootstrap

Layout, Nav-Tabs, Modal, Icons

#### SCSS

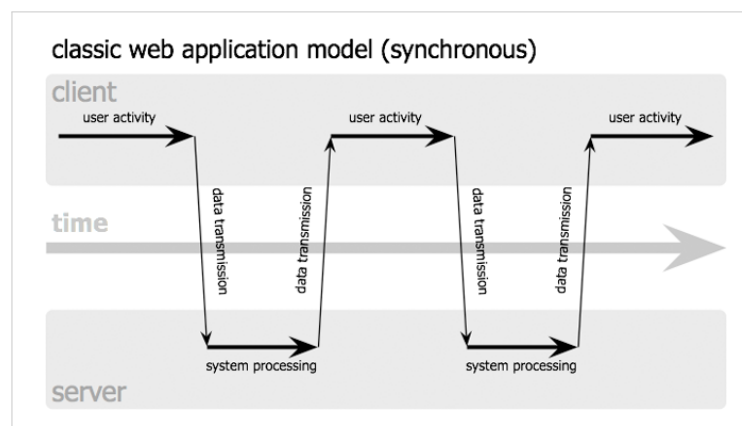
SASS (Catlin, Weizenbaum und Eppstein, o. D.): Einfachere Syntax, Variablen, Funktionen

#### Tooltips

Popper (Popper JS, o. D.)

### 4.5 AJAX

AJAX (Garrett, 2005)





## 4 Web-Applikation

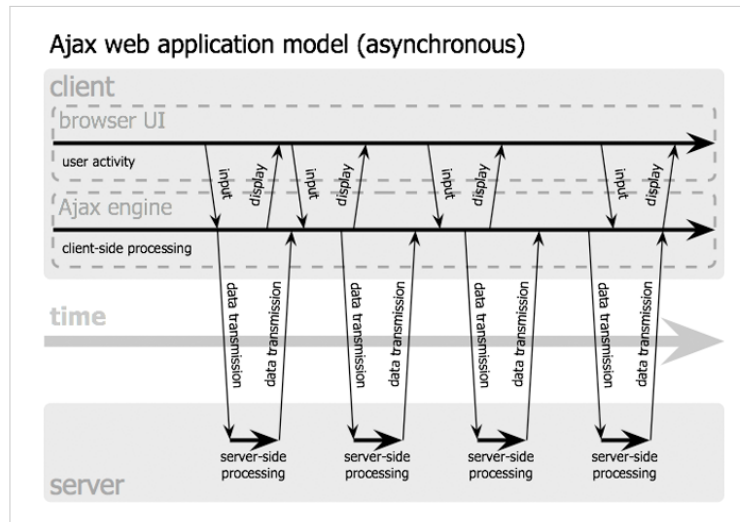


Abbildung 4.2: (Garrett, 2005)

Modal

BfArM/Dimdi Links – Neuer Tab muss geöffnet werden bevor Skripte ausgeführt werden.

CSS Flexbox (Coyier, o. D.)

### Anonyme Funktionen

Callback

Zum Schreiben der ConceptMap

Data-Preprocessing

InsertEncoder

<https://dev.mysql.com/doc/refman/en/insert-optimization.html>

# 5

## Zusammenfassung

### 5.1 Ergebnisse

#### Gegenseitige Validierung / Intra-DB Validierung

(Thurin u. a., 2021)

#### Patienten-DB

FHIR ConceptMap curl (Stenberg u. a., o. D.)

bfarmer			
Test-Datenbank ICD-10-GM			
Patienten insgesamt: 20000			
Code:	Titel:	Lähmungssyndrome	Suchen 4 Treffer
Patient	Jahr	Code	Titel
10408	2023	G83.8	Sonstige näher bezeichnete Lähmungssyndrome
		K58.8	Sonstiges und nicht näher bezeichnetes Reizdarmsyndrom
		S42.44	Fraktur des distalen Endes des Humerus: Epicondylus, Epicondylus, nicht näher bezeichnet
05248	2020	G83.8	Sonstige näher bezeichnete Lähmungssyndrome
		K41.30	Hernia femoralis, einseitig oder ohne Seitenangabe, mit Einklemmung, ohne Gangrän: Nicht als Rezidivhernie bezeichnet
		M12.56	Traumatische Arthropathie: Unterschenkel [Fibula, Tibia, Kniegelenk]
		S83.43	Verstauchung und Zerrung des Kniegelenkes: Riss des fibularen Seitenbandes [Außenband]
03611	2013	G83.88	Sonstige näher bezeichnete Lähmungssyndrome
		H40.5	Glaukom (sekundär) nach sonstigen Affektionen des Auges
00352	1.3	G83.8	Sonstige näher bezeichnete Lähmungssyndrome

Abbildung 5.1: Patienten

## 5 Zusammenfassung

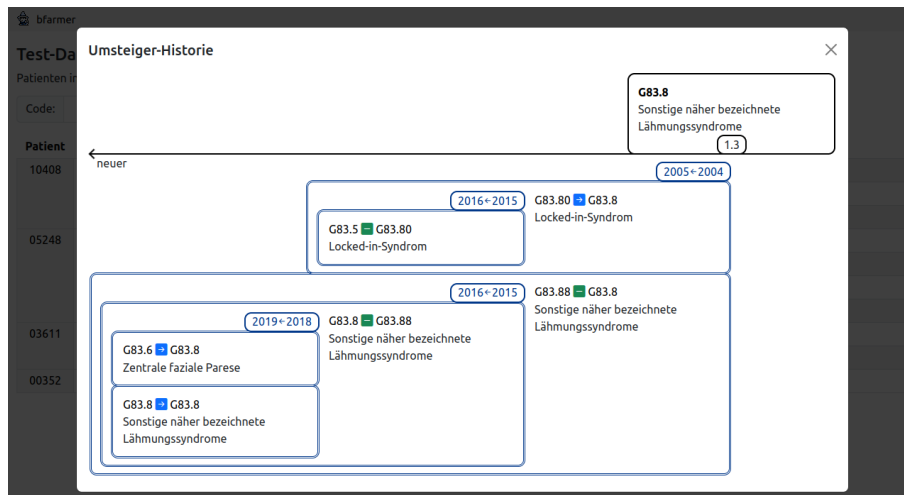


Abbildung 5.2: Umsteiger

Integration:

1. Bootstrap
2. CSS-Datei
3. On-Click Event
4. Ajax-Funktion

### Mapping Quality

Beurteilung der generierten ConceptMaps anhand der Qualitätsstandards aus Abschnitt 1.2:

## 5.2 Zusammenfassung

## 5.3 Ausblick

Folgende Weiterentwicklungen sind sinnvoll.

### RESTful

Die Web-Applikation sollte nicht nur teilweise, sondern komplett dem REST-Paradigma entsprechen, wie in Abschnitt 1.1 aufgeführt.

Das bedeutet alle Informationen des Servers durch Schnittstellen bereitzustellen:

- Kodes als FHIR Code System
- Umsteiger pro Version im JSON-Format, AJAX für Übersichtseite

- Ergebnis der horizontalen Suche im JSON-Format
- ConceptMap ohne Formular

Außerdem sollten die Schnittstellen dokumentiert sein über Swagger UI (Weaver, 2023).

Bei einem wirklich produktiv eingesetzten System wäre es zudem sinnvoll folgende Dinge zu cachen:

- Umsteiger-Ergebnisse
- ConceptMaps

Das würde allerdings sehr viel mehr an Speicherplatz benötigen.

### ATC

ATC (BfArM, o. D.[a], ATC-Klassifikation)

PDF (Poppler, o. D.)

### Sessions

## 5 Zusammenfassung

---

```
curl -X 'POST' \  
  'http://localhost:8080/fhir/ConceptMap' \  
  -H 'accept: application/fhir+json' \  
  -H 'Content-Type: application/fhir+xml' \  
  -d "@/home/simon/Downloads/conceptmap_r4_icd10gm_2024.xml"
```

---

---

```
curl -X 'GET' \  
  'http://localhost:8080/fhir/ConceptMap/$translate?code=G83.8&system=1.3' \  
  -H 'accept: application/fhir+json'
```

---

# Literatur

Adermann, N., Boggiano, J. u. a. (o. D.). *Composer. A Dependency Manager for PHP*. URL: <https://getcomposer.org/doc/> (besucht am 12. 08. 2024).

BfArM (o. D.[a]). *BfArM – ATC. ATC-Klassifikation*. URL: [https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ATC/\\_node.html](https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ATC/_node.html) (besucht am 23. 10. 2024).

BfArM (o. D.[b]). *BfArM – Downloads*. URL: [https://www.bfarm.de/DE/Kodiersysteme/Services/Downloads/\\_node.html](https://www.bfarm.de/DE/Kodiersysteme/Services/Downloads/_node.html) (besucht am 09. 09. 2024).

BfArM (o. D.[c]). *BfArM – Geschichte – Geschichte des ehemaligen DIMDI. Geschichte des ehemaligen DIMDI*. URL: <https://www.bfarm.de/DE/Das-BfArM/Organisation/Geschichte/geschichte-dimdi.html> (besucht am 25. 10. 2024).

BfArM (o. D.[d]). *BfArM – Systematisches Verzeichnis – Kategorie und Kode. Kategorie und Kode in der ICD-10-GM*. URL: <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/ICD/ICD-10-GM/Systematik/kodestruktur.html> (besucht am 09. 09. 2024).

BfArM (o. D.[e]). *BfArM – Systematisches Verzeichnis – Kategorie und Kode. Kategorie und Kode im OPS*. URL: <https://www.bfarm.de/DE/Kodiersysteme/Klassifikationen/OPS-ICHI/OPS/Systematik/kategorie-und-kode.html> (besucht am 09. 09. 2024).

Bonnefoy, P., Chaize, E., Mansuy, R., Tazi, M. und Heckel, S. (2024). *The Definitive Guide to Data Integration: Unlock the Power of Data Integration to Efficiently Manage, Transform, and Analyze Data*. Packt Publishing. ISBN: 9781837634774.

Braaksma, A. (2014). *Streaming Design Patterns or: How I Learned to Stop Worrying and Love the Stream*. XML LONDON 2014, 24–52. URL: <http://xmllondon.com/2014/xmllondon-2014-proceedings.pdf> (besucht am 22. 09. 2024).

Braunstein, M. (2022). *Health Informatics on FHIR: How HL7's API is Transforming Healthcare*. Health Informatics. Springer International Publishing. ISBN: 9783030915636.

Catlin, H., Weizenbaum, N. und Eppstein, C. (o. D.). *SASS*. URL: <https://sass-lang.com/documentation/> (besucht am 12. 08. 2024).

Coyier, C. (o. D.). *CSS Flexbox Layout Guide*. URL: <https://css-tricks.com/snippets/css/a-guide-to-flexbox/> (besucht am 09. 09. 2024).

- Dahl, R. (o. D.). *Node.js*. URL: <https://nodejs.org/docs/latest/api/> (besucht am 12.08.2024).
- Dar, S. (1993). *Augmenting Databases with Generalized Transitive Closure*. Computer Sciences Technical Report. University of Wisconsin–Madison.
- Ecma International (o. D.). *ECMAScript. JavaScript*. URL: <https://ecma-international.org/publications-and-standards/standards/ecma-262/> (besucht am 12.08.2024).
- Fernández, I. und Manuel, J. (2022). *UTF-8 & Latex Encodings of ISO-8859 (Latin-1) Character Set*. DOI: 10.13140/RG.2.2.18402.61121. URL: [https://www.researchgate.net/publication/359509972\\_UTF-8\\_Latex\\_Encodings\\_of\\_ISO-8859\\_Latin-1\\_Character\\_Set/link/6241ab3b5e2f8c7a03452ac9/download](https://www.researchgate.net/publication/359509972_UTF-8_Latex_Encodings_of_ISO-8859_Latin-1_Character_Set/link/6241ab3b5e2f8c7a03452ac9/download) (besucht am 06.09.2024).
- Fielding, R.T. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral Dissertation. University of California, Irvine. URL: <https://ics.uci.edu/~fielding/pubs/dissertation/top.htm> (besucht am 04.10.2024).
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M. und Herrera, F. (2016). Big Data Preprocessing: Methods and Prospects. *Big Data Analytics* 1, 1–22. DOI: 10.1186/s41044-016-0014-0. URL: <https://bdataanalytics.biomedcentral.com/articles/10.1186/s41044-016-0014-0>.
- Garrett, J.J. (2005). *Ajax: A New Approach to Web Applications*. URL: <https://web.archive.org/web/20080702075113/http://www.adaptivepath.com/ideas/essays/archives/000385.php> (besucht am 14.10.2024).
- Gaus, W. (2005). *Dokumentations- und Ordnungslehre*. Springer. ISBN: 9783540275183.
- Gross, J., Yellen, J. und Zhang, P. (2013). *Handbook of Graph Theory, Second Edition. Discrete Mathematics and Its Applications*. Taylor & Francis. ISBN: 9781439880180.
- Heckmann, S. (2022). *Digitalstrategie im Krankenhaus: Einführung und Umsetzung von Datenkompetenz und Compliance*. Springer Fachmedien Wiesbaden. Kap. HL7® FHIR® als Grundlage für moderne Digitalstrategien, S. 307–331. ISBN: 9783658362263. DOI: 10.1007/978-3-658-36226-3\_21.
- HL7 (2019). *ConceptMap of FHIR Specification R4. Resource ConceptMap*. URL: <http://hl7.org/fhir/R4/conceptmap.html> (besucht am 06.06.2024).
- HL7 (2023). *ConceptMap of FHIR Specification R5. Resource ConceptMap*. URL: <http://hl7.org/fhir/R5/conceptmap.html> (besucht am 20.06.2024).
- Hund, H. (2018). *medicats. Medical classification and terminology systems library for Java 1.8*. URL: <https://github.com/hhund/medicats> (besucht am 15.05.2024).

Hund, H., Gerth, S., Katus, H. und Fegeler, C. (2016). Medical Classification and Terminology Systems in a Secondary Use Context: Challenges and Perils. *Studies in health technology and informatics* 228, 394–398.

ISO (2014). *ISO/TR 12300:2014 – Principles of Mapping between Terminological Systems*. International Organization for Standardization, Technical Committee: Health Informatics. URL: <https://www.iso.org/standard/51344.html> (besucht am 21. 10. 2024).

ISO (2019). *ISO/TS 21564:2019 – Terminology Resource Map Quality Measures (MapQual)*. International Organization for Standardization, Technical Committee: Health Informatics. URL: <https://www.iso.org/standard/71088.html> (besucht am 21. 10. 2024).

Jakobsson, H. (1991). Mixed-Approach Algorithms for Transitive Closure. In *Proceedings of the Tenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, S. 199–205. URL: <https://dl.acm.org/doi/pdf/10.1145/113413.113431>.

Janssens, S., Schultz, U.P. und Zaytsev, V. (2017). *Can Some Programming Languages Be Considered Harmful?* URL: <https://grammarware.net/text/2017/harmful.pdf> (besucht am 10. 10. 2024).

Koppers, T. u. a. (o. D.). *Webpack*. URL: <https://webpack.js.org/concepts/> (besucht am 12. 08. 2024).

Lawley, M. (2023). *Terminology: ConceptMap “dependsOn”, “product”, and “unmapped” Relationships*. URL: <https://chat.fhir.org/#narrow/stream/179202-terminology/topic/conceptmap.20dependson.20product.20and.20unmapped.20relationships/near/326784399> (besucht am 16. 09. 2024).

Medizininformatik-Initiative (o. D.). *Forschungsdatenportal für Gesundheit*. URL: <https://forschen-fuer-gesundheit.de> (besucht am 02. 07. 2024).

Nassi, I. und Shneiderman, B. (Aug. 1973). Flowchart Techniques for Structured Programming. *ACM SIGPLAN Notices* 8, 12–26. DOI: 10.1145/953349.953350. URL: [https://www.researchgate.net/publication/234805404\\_Flowchart\\_techniques\\_for\\_structured\\_programming](https://www.researchgate.net/publication/234805404_Flowchart_techniques_for_structured_programming).

Philipp, P., Oliveira, J. Veloso de, Appenzeller, A., Hartz, T. und Beyerer, J. (Dez. 2022). Evaluation of an Automated Mapping from ICD-10 to SNOMED CT. In *2022 International Conference on Computational Science and Computational Intelligence*, S. 1604–1609. DOI: 10.1109/CSCI58124.2022.00287.

Popper JS (o. D.). *Tooltip & Popover Positioning Engine*. URL: <https://popper.js.org/docs/v2/> (besucht am 12. 08. 2024).

Poppler (o. D.). *Poppler is a PDF rendering library based on the xpdf-3.0 code base*. URL: <https://poppler.freedesktop.org/> (besucht am 12. 08. 2024).



- Potencier, F. (2022). *Symfony 6: The Fast Track*. Symfony SAS. URL: <https://symfony.com/doc/6.4/the-fast-track/en/index.html> (besucht am 22. 09. 2024).
- Purdom, P. (1970). A Transitive Closure Algorithm. *BIT Numerical Mathematics* 10, 76–94.
- Saripalle, R. (2019). Representing UMLS knowledge using FHIR terminological resources. In 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, S. 1109–1112. DOI: 10.1109/BIBM47256.2019.8983305.
- Schlueter, I.Z. (o. D.). *npm*. URL: <https://docs.npmjs.com/> (besucht am 12. 08. 2024).
- Seemann, M. und Deursen, S. van (2019). *Dependency Injection Principles, Practices, and Patterns*. Manning. ISBN: 9781638357100.
- Stenberg, D. u. a. (o. D.). *cURL*. URL: <https://curl.se/docs/> (besucht am 04. 10. 2024).
- Symfony (o. D.[a]). *Symfony 6.4. Console Commands*. URL: <https://symfony.com/doc/6.4/console.html> (besucht am 12. 08. 2024).
- Symfony (o. D.[b]). *Symfony 6.4. The HttpFoundation Component*. URL: [https://symfony.com/doc/6.4/components/http\\_foundation.html](https://symfony.com/doc/6.4/components/http_foundation.html) (besucht am 12. 08. 2024).
- Symfony (o. D.[c]). *Symfony 6.4. HTTP Client*. URL: [https://symfony.com/doc/6.4/http\\_client.html](https://symfony.com/doc/6.4/http_client.html) (besucht am 12. 08. 2024).
- Symfony (o. D.[d]). *Symfony 6.4. The Filesystem Component*. URL: <https://symfony.com/doc/6.4/components/filesystem.html> (besucht am 12. 08. 2024).
- Symfony (o. D.[e]). *Symfony 6.4. The UID Component*. URL: <https://symfony.com/doc/6.4/components/uid.html> (besucht am 12. 08. 2024).
- Symfony (o. D.[f]). *Symfony 6.4. Controller*. URL: <https://symfony.com/doc/6.4/controller.html> (besucht am 12. 08. 2024).
- Symfony (o. D.[g]). *Symfony 6.4. The Serializer Component*. URL: <https://symfony.com/doc/6.4/components/serializer.html> (besucht am 12. 08. 2024).
- Symfony (o. D.[h]). *Symfony 6.4. Databases and the Doctrine ORM*. URL: <https://symfony.com/doc/6.4/doctrine.html> (besucht am 12. 08. 2024).
- Symfony (o. D.[i]). *Symfony 6.4. Creating and Using Templates*. URL: <https://symfony.com/doc/6.4/templates.html> (besucht am 12. 08. 2024).
- Thomas, D. und Hunt, A. (2019). *The Pragmatic Programmer: Your Journey to Mastery*. Addison Wesley. ISBN: 9780135957059.
- Thurin, N., Bosco, P., Blin, P., Rouyer, M., Jové, J., Lamarque, S., Lignot, S., Lassalle, R., Abouelfath, A., Bignon, E. u. a. (2021). Intra-Database Validation of Case-Identifying Al-

gorithms using Reconstituted Electronic Health Records from Healthcare Claims Data. BMC Medical Research Methodology 21. DOI: 10.1186/s12874-021-01285-y.

Twig (o.D.). *Twig 3.x. The flexible, fast, and secure template engine for PHP*. URL: <https://twig.symfony.com/doc/3.x/> (besucht am 12.08.2024).

Voorhees, D. (2020). *Guide to Efficient Software Design: An MVC Approach to Concepts, Structures, and Models*. Texts in Computer Science. Springer International Publishing. ISBN: 9783030285012.

W3C (o.D.). *CSS*. URL: <https://www.w3.org/TR/CSS/> (besucht am 12.08.2024).

W3Techs (o.D.). *Usage Statistics of PHP for Websites*. URL: <https://w3techs.com/technologies/details/pl-php> (besucht am 25.10.2024).

Weaver, R. (2023). *Swagger UI: Interactive Docs*. URL: <https://symfonycasts.com/screencast/api-platform/swagger> (besucht am 25.10.2024).

WHATWG (o.D.). *HTML*. URL: <https://html.spec.whatwg.org/> (besucht am 12.08.2024).

WHO-FIC (2021). *Classifications and Terminology Mapping – Principles and Best Practice*. World Health Organization, Group: Family of International Classifications. URL: <https://www.who.int/publications/m/item/who-fic-classifications-andterminology-mapping> (besucht am 21.10.2024).



## Appendix

### A.1 Abweichungen zwischen ICD-10-GM und OPS Versionen

#### ICD-10-GM

Version	Abweichungen zwischen den Versionen	
2025	URL	version2025-vorab/icd10gm2025syst-ueberl-vorab.zip
	Kodes	Klassifikationsdateien/icd10gm2025syst_vorab.txt
	Umsteiger	Klassifikationsdateien/ icd10gm2025syst_umsteiger_2024_2025_vorab.txt
	Sonstiges	· Vorab-Version
2024	URL	version2024/icd10gm2024syst-ueberl.zip
	Umsteiger	Klassifikationsdateien/ icd10gm2024syst_umsteiger_2023_20221206_2024.txt
2023	URL	version2023/icd10gm2023syst-ueberl_20221206.zip
	Kodes	Klassifikationsdateien/ icd10gm2023syst_20221206.txt
	Umsteiger	Klassifikationsdateien/ icd10gm2023syst_umsteiger_2022_2023_20221206.txt
2022	URL	vorgaenger/icd10gm2022.zip
	Sonstiges	· Zip-Unterdatei: icd10gm2022syst-ueberl.zip
2021	URL	vorgaenger/icd10gm2021.zip
	Verzeichnis	icd10gm2021syst-ueberl-20201111
2020	URL	vorgaenger/icd10gm2020.zip
	Verzeichnis	icd10gm2020syst-ueberl
2019	URL	vorgaenger/icd10gm2019.zip
	Verzeichnis	icd10gm2019syst-ueberl

## A Appendix

2018	URL	vorgaenger/icd10gm2018.zip
	Verzeichnis	x1gut2018
2017	URL	vorgaenger/icd10gm2017.zip
	Verzeichnis	x1gut2017
2016	URL	vorgaenger/icd10gm2016.zip
	Verzeichnis	x1gut2016
2015	URL	vorgaenger/icd10gm2015.zip
	Verzeichnis	x1gut2015
2014	URL	vorgaenger/icd10gm2014.zip
	Verzeichnis	x1gua2014
2013	URL	vorgaenger/icd10gm2013.zip
	Verzeichnis	x1gua2013
2012	URL	vorgaenger/icd10gm2012.zip
	Kodes	x1ueb2011_2012/Klassifikationsdateien/ icd10gmsyst2012.txt
	Umsteiger	x1ueb2011_2012/Klassifikationsdateien/ umsteiger_icd10gmsyst2011_icd10gmsyst2012.txt
2011	URL	vorgaenger/icd10gm2011.zip
	Kodes	x1ueb2010_2011/Klassifikationsdateien/ icd10gmsyst2011.txt
	Umsteiger	x1ueb2010_2011/Klassifikationsdateien/ umsteiger_icd10gmsyst2010_icd10gmsyst2011.txt
2010	URL	vorgaenger/icd10gm2010.zip
	Kodes	x1ueb2009_2010/Klassifikationsdateien/ icd10gmsyst2010.txt
	Umsteiger	x1ueb2009_2010/Klassifikationsdateien/ umsteiger_icd10gmsyst2009_icd10gmsyst2010.txt
2009	URL	vorgaenger/icd10gm2009.zip
	Kodes	x1ueb2008_2009/Klassifikationsdateien/ icd10gmsyst2009.txt
	Umsteiger	x1ueb2008_2009/Klassifikationsdateien/ umsteiger_icd10gmsyst2008_icd10gmsyst2009.txt
2008	URL	vorgaenger/icd10gm2008.zip
	Kodes	x1ueb2007_2008/Klassifikationsdateien/ icd10v2008.txt
	Umsteiger	x1ueb2007_2008/Klassifikationsdateien/ umsteiger20072008.txt
	Sonstiges	· ISO-8859-1

## A Appendix

2007	URL	vorgaenger/icd10gm2007.zip
	Kodes	x1ueb2006_2007/Klassifikationsdateien/ ICD10V2007.txt
	Umsteiger	x1ueb2006_2007/Klassifikationsdateien/ Umsteiger.txt
	Sonstiges	· ISO-8859-1
2006	URL	vorgaenger/icd10gm2006.zip
	Kodes	x1ueb2005_2006/ICD10V2006.txt
	Umsteiger	x1ueb2005_2006/umsteiger.txt
	Sonstiges	· ISO-8859-1
2005	URL	vorgaenger/icd10gm2005.zip
	Kodes	x1ueb2004_2005/ICD10V2005.txt
	Umsteiger	x1ueb2004_2005/umsteiger.txt
	Sonstiges	· ISO-8859-1
2004	URL	vorgaenger/icd10gm2004.zip
	Kodes	x1ueb20_2004/icd10v2004.txt
	Umsteiger	x1ueb20_2004/Umsteiger.txt
	Sonstiges	· ISO-8859-1 · Punkt-Strich-Notation · 6-Spalten-Umsteiger
2.0	URL	vorgaenger/icd10gm20.zip
	Kodes	x1ueb13_20_v11/icd10v20.txt
	Umsteiger	x1ueb13_20_v11/Umsteiger.txt
	Sonstiges	· ISO-8859-1 · Punkt-Strich-Notation · Kreuz-Stern-System · 6-Spalten-Umsteiger · Nicht endständige Umsteiger
1.3	Kodes	x1ueb13_20_v11/icd10v13.txt
	Sonstiges	· ISO-8859-1 · Punkt-Strich-Notation · Kreuz-Stern-System · Keine Überleitung

## OPS

Version	Abweichungen zwischen den Versionen	
2025	URL	version2025-vorab/ops2025syst-ueberl-vorab.zip
	Kodes	Klassifikationsdateien/ops2025syst_vorab.txt
	Umsteiger	Klassifikationsdateien/ ops2025syst_umsteiger_2024_2025_vorab.txt
	Sonstiges	· Vorab-Version
2024	URL	version2024/ops2024syst-ueberl.zip
2023	URL	version2023/ops2023syst-ueberl.zip
2022	URL	vorgaenger/ops2022.zip
	Sonstiges	· Zip-Unterdatei: ops2022syst-ueberl.zip
2021	URL	vorgaenger/ops2021.zip
	Verzeichnis	ops2021syst-ueberl
2020	URL	vorgaenger/ops2020.zip
	Verzeichnis	ops2020syst-ueberl
2019	URL	vorgaenger/ops2019.zip
	Verzeichnis	ops2019syst-ueberl
2018	URL	vorgaenger/ops2018.zip
	Verzeichnis	p1sut2018
2017	URL	vorgaenger/ops2017.zip
	Verzeichnis	p1sut2017
2016	URL	vorgaenger/ops2016.zip
	Verzeichnis	p1sut2016
2015	URL	vorgaenger/ops2015.zip
	Verzeichnis	p1sut2015
2014	URL	vorgaenger/ops2014.zip
	Kodes	p1sua2014-20131104/Klassifikationsdateien/ ops2014syst_20131104.txt
	Umsteiger	p1sua2014-20131104/Klassifikationsdateien/ ops2014syst_umsteiger_2013_2014_20131104.txt
2013	URL	vorgaenger/ops2013.zip
	Kodes	p1sua2013/Klassifikationsdateien/ ops2013syst_20121113.txt
	Umsteiger	p1sua2013/Klassifikationsdateien/ ops2013syst_umsteiger_2012_2013_20121109.txt

## A Appendix

2012	URL	vorgaenger/ops2012.zip
	Kodes	p1ueb2011_2012/Klassifikationsdateien/ opssyst2012.txt
	Umsteiger	p1ueb2011_2012/Klassifikationsdateien/ umsteiger_opssyst2011_opssyst2012.txt
2011	URL	vorgaenger/ops2011.zip
	Kodes	p1ueb2010_2011/Klassifikationsdateien/ opssyst2011.txt
	Umsteiger	p1ueb2010_2011/Klassifikationsdateien/ umsteiger_opssyst2010_opssyst2011.txt
2010	URL	vorgaenger/ops2010.zip
	Kodes	p1ueb2009_2010/Klassifikationsdateien/ opssyst2010.txt
	Umsteiger	p1ueb2009_2010/Klassifikationsdateien/ umsteiger_opssyst2009_opssyst2010.txt
2009	URL	vorgaenger/ops2009.zip
	Kodes	p1ueb2008_2009/Klassifikationsdateien/ opsamtl2009.txt
	Umsteiger	p1ueb2008_2009/Klassifikationsdateien/ umsteigeramtl20082009.txt
	Sonstiges	· ISO-8859-1 · None statt UNDEF · 6-Spalten-Umsteiger (altes Format)
2008	URL	vorgaenger/ops2008.zip
	Kodes	ops2008amtl/p1ueb2007_2008/ Klassifikationsdateien/opsamtl2008.txt
	Umsteiger	ops2008amtl/p1ueb2007_2008/ Klassifikationsdateien/umsteigeramtl20072008.txt
	Sonstiges	· ISO-8859-1 · None statt UNDEF · 6-Spalten-Umsteiger (altes Format)
2007	URL	vorgaenger/ops2007.zip
	Kodes	ops2007amtl/p1ueb2006_2007/ Klassifikationsdateien/opsamtl2007.txt
	Umsteiger	ops2007amtl/p1ueb2006_2007/ Klassifikationsdateien/UmsteigerAmtlich.txt
	Sonstiges	· ISO-8859-1 · None statt UNDEF · 6-Spalten-Umsteiger (altes Format)

## A Appendix

2006	URL	vorgaenger/ops2006.zip
	Kodes	ops2006amtl/p1ueb2005_2006/opsv2006.txt
	Umsteiger	ops2006amtl/p1ueb2005_2006/umsteiger.txt
	Sonstiges	· ISO-8859-1 · None statt UNDEF · 6-Spalten-Umsteiger (altes Format)
2005	URL	vorgaenger/ops2005.zip
	Kodes	ops2005amtl/p1ueb2004_2005_v10/OPS2005.txt
	Umsteiger	ops2005amtl/p1ueb2004_2005_v10/umsteiger.txt
	Sonstiges	· ISO-8859-1 · None statt UNDEF · 5-Spalten-Umsteiger
2004	URL	vorgaenger/ops2004.zip
	Kodes	ops2004amtl/p1ueb21_2004_v10/opsv2004.txt
	Umsteiger	ops2004amtl/p1ueb21_2004_v10/Umsteiger.txt
	Sonstiges	· ISO-8859-1 · 4-Spalten-Umsteiger
2.1	URL	vorgaenger/ops21.zip
	Kodes	ops21amtl/p1ueb20_21_v10/opsv21.txt
	Umsteiger	ops21amtl/p1ueb20_21_v10/Umsteiger.txt
	Sonstiges	· ISO-8859-1 · KOMBI-Kode · 6-Spalten-Umsteiger (ursprüngliches Format)
2.0	URL	vorgaenger/ops20.zip
	Kodes	p1ueb11_20_v11/0psv20.txt
	Umsteiger	p1ueb11_20_v11/Umsteiger.txt
	Sonstiges	· ISO-8859-1 · KOMBI-Kode · 3-Spalten-Umsteiger
1.1	Kodes	p1ueb11_20_v11/0psv11.txt
	Sonstiges	· ISO-8859-1 · Keine Überleitung



## A.2 Beispielergebnis der Horizontalen Suche

Ergebnis im JSON-Format für den Aufruf:

```
searchHorizontal('icd10gm', '2014', 'M21.6')
```

Um Platz zu sparen enthält nur der erste Umsteiger die zusätzlichen Informationen, die pro Eintrag ermittelt werden, also die automatische Überleitbarkeit sowie die Titel der Kodes. Außerdem sind für die Lesbarkeit die Sonderzeichen nicht kodiert.

Die Seite zeigt die Richtung 2024←2014 und die nächste Seite 2014→2004, 2.0, 1.3.

---

```
{
  "fwd": {
    "year": "2014",
    "other": "2015",
    "umsteiger": [
      {
        "old": "M21.6",
        "new": "M21.60",
        "auto": "",
        "auto_r": "A",
        "old_name": "Sonstige erworbene Deformitäten des Knöchels und des Fußes",
        "new_name": "Erworbener Hohlfuß [Pes cavus]"
      },
      {
        "old": "M21.6",
        "new": "M21.61",
      },
      {
        "old": "M21.6",
        "new": "M21.62",
      },
      {
        "old": "M21.6",
        "new": "M21.63",
      },
      {
        "old": "M21.6",
        "new": "M21.68",
      }
    ]
  },
}
```

```

"rev": {
  "year": "2013",
  "other": "2012",
  "umsteiger": [
    {
      "old": "M21.60",
      "new": "M21.6",
      "recursion": {
        "year": "2.0",
        "other": "1.3",
        "umsteiger": [
          {
            "old": "M21.6",
            "new": "M21.60",
          }
        ]
      }
    }
  ],
},
{
  "old": "M21.67",
  "new": "M21.6",
  "recursion": {
    "year": "2.0",
    "other": "1.3",
    "umsteiger": [
      {
        "old": "M21.6",
        "new": "M21.67",
      }
    ]
  }
},
{
  "old": "M21.87",
  "new": "M21.6",
  "recursion": {
    "year": "2.0",
    "other": "1.3",
    "umsteiger": [
      {
        "old": "M21.8",
        "new": "M21.87",
      }
    ]
  }
}
]
}
}

```

---

### A.3 Beispielergebnis der Vertikalen Suche / ConceptMap

FHIR ConceptMap im XML-Format für das Beispiel in Abbildung 3.6. Die komplette ConceptMap enthält noch viel mehr Gruppen und Codes, aber abgedruckt sind nur die relevanten Teile wie beschrieben in Unterabschnitt 3.2 – „Konkretes Beispiel“.

---

```

<?xml version="1.0"?>
<ConceptMap xmlns="http://hl7.org/fhir">
  <id value="icd10gm_from:2024_to:1.3_target:2024"/>
  <url value="urn:uuid:0192e098-0871-79d5-a4e6-039398aa7803"/>

  <group>
    <source value="2018"/>
    <target value="2024"/>
    <element>
      <code value="G83.8"/>
      <target>
        <code value="G83.6"/>
        <equivalence value="wider"/>
      </target>
      <target>
        <code value="G83.8"/>
        <equivalence value="wider"/>
      </target>
    </element>
  </group>

  <group>
    <source value="2015"/>
    <target value="2024"/>
    <element>
      <code value="G83.80"/>
      <target>
        <code value="G83.5"/>
        <equivalence value="relatedto"/>
      </target>
    </element>
    <element>
      <code value="G83.88"/>
      <target>
        <code value="G83.6"/>
        <equivalence value="wider"/>
      </target>
      <target>
        <code value="G83.8"/>
        <equivalence value="wider"/>
      </target>
    </element>
  </group>

```

```

<group>
  <source value="2004"/>
  <target value="2024"/>
  <element>
    <code value="G83.8"/>
    <target>
      <code value="G83.6"/>
      <equivalence value="wider"/>
    </target>
    <target>
      <code value="G83.8"/>
      <equivalence value="wider"/>
    </target>
    <target>
      <code value="G83.5"/>
      <equivalence value="wider"/>
    </target>
  </element>
</group>

<group>
  <source value="1.3"/>
  <target value="2024"/>
  <element>
    <code value="G83.8"/>
    <target>
      <code value="G83.6"/>
      <equivalence value="wider"/>
    </target>
    <target>
      <code value="G83.8"/>
      <equivalence value="wider"/>
    </target>
    <target>
      <code value="G83.5"/>
      <equivalence value="wider"/>
    </target>
  </element>
</group>
</ConceptMap>

```

---