



hochschule mannheim

in Kooperation mit



miracum

BIOMEDIZINISCHE INFORMATIK UND DATA SCIENCE
Master of Science (M.Sc.)

MASTERARBEIT

Wissenschaftliche Nachnutzung der Biosignaldaten aus der Routineversorgung in einem deutschen Universitätsklinikum

Abel HODELÍN HERNÁNDEZ 

betreut von
Daniel SCHMITZ

18. November 2022



Universitätsklinikum
Erlangen



Philipps
Universität
Marburg



FACHBEREICH
MEDIZIN FRANKFURT
GOETHE-UNIVERSITÄT



UKGM
UNIVERSITÄTSKLINIKUM
GIESSEN UND MARBURG



TECHNISCHE
UNIVERSITÄT
DRESDEN



THM
TECHNISCHE HOCHSCHULE MITTELHESSEN



hochschule mannheim



UNIVERSITÄTS
KLINIKUM
ERLANGEN



JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN



averbis
text analytics



Universitätsklinikum
Carl Gustav Carus



UNIVERSITÄT
MAGDEBURG



Medizinische Fakultät
UNIVERSITÄTSKLINIKUM MAGDEBURG A.G.R.



UMM
UNIVERSITÄTSMEDIZIN
MANNHEIM



Medizinische Fakultät
der Universität Mannheim
Universitätsklinikum Mannheim



Universitätsmedizin
GRIFFSALD



Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung	1
1.2	Ziel der Arbeit	2
1.3	Aufbau der Arbeit	3
2	Theoretische Grundlagen	5
2.1	Interoperabilität	5
2.1.1	Interoperabilität im Gesundheitswesen	6
2.2	Kerndatensatz der MII	7
2.2.1	Erweiterungsmodul „Intensivmedizin“ des Kerndaten- satzes der MII	8
2.3	HL7 & FHIR	11
2.3.1	HL7	11
2.3.2	FHIR	12
2.4	Codesysteme	17
2.4.1	SNOMED CT	18
2.4.2	LOINC	19
2.4.3	ISO/IEEE 11073™	21
2.4.4	UCUM	22
2.5	Patientendatenmanagementsystem	22
2.5.1	COPRA	22
2.6	Data Warehouse	24
2.6.1	ETL	25
2.6.2	Datenmodell eines DWH	26
2.7	Data Mapping	27
2.8	Pattern Matching	29
2.8.1	Unscharfe Suche	30
2.8.2	Regex	30

3	Methode	32
3.1	Schritte für die Durchführung	32
3.2	Data Mapping der Biosignaldaten aus COPRA mit den FHIR- Profilen	34
3.2.1	Datendefinition	35
3.2.2	Zuweisung der Quell- und Zielfelder	35
3.2.3	Transformationsregeln	36
3.3	Analyse für die Harmonisierung der Maßeinheiten	36
3.4	Vorbereitung der Überführung der Daten aus COPRA in FHIR	37
4	Realisierung	38
4.1	PostgreSQL-DB	39
4.2	Parameter der FHIR-Profile des Moduls „ICU“	40
4.3	Analyse der FHIR-Profile des MII - Moduls ICU	43
4.4	Angepasste Tabellen von COPRA	45
4.4.1	Werttabellen	47
4.5	Auswahl der Konfigurationsvariablen von COPRA	50
4.6	Analyse der Tabellen für die Datendefinition	51
4.7	Pattern Matching der FHIR-Profile mit den Konfigurations- variablen	57
4.8	Validierung des Pattern Matchings	61
4.9	Harmonisierung der Maßeinheiten	62
4.10	Zuweisung der Quell- und Zielfelder und weitere Transforma- tionsregeln	65
4.10.1	Modellierung der Zuweisung der Felder beider Systeme	66
4.11	Bereitstellung der Biosignaldaten	71
5	Ergebnisse und Diskussion	77
5.1	Zuordnung der COPRA-Biosignaldaten mit den FHIR-Profilen des Moduls „Intensivmedizin“	77
5.2	FHIR-Profile	79
5.3	Biosignaldaten aus COPRA	81
6	Fazit und Ausblick	84
	Literatur	85
A	SQL	92
B	URI, URL und URN	93
C	Flussdiagramm	94

Abbildungsverzeichnis	95
Tabellenverzeichnis	96
Listingverzeichnis	98
Abkürzungsverzeichnis	99
Danksagung	101

Zusammenfassung

Der Prozess der Digitalisierung im Gesundheitswesen ist ein wichtiger Bestandteil eines sozialen Wandlungsprozesses. Eine der zentralen Herausforderungen in diesem Prozess ist die mangelnde Interoperabilität vieler Systeme, zusammen mit dem gewaltigen generierten Datenvolumen während der Routineversorgung, denn die erfassten Parameter werden nicht immer digital erfasst und werden auch in diversen getrennten Systemen gesteuert.

Ziel dieser Arbeit ist die Bereitstellung der in einem Patientendatenmanagementsystem (PDMS) gespeicherten Biosignaldaten aus der Routineversorgung eines deutschen Universitätsklinikums für die Überführung dieser Bioparameter in das etablierte Standardformat Fast Healthcare Interoperability Resources (FHIR), sodass die Interoperabilität der Biosignaldaten gewährleistet wird, und diese Daten in der Zukunft integriert und wieder anwendbar werden können. Dazu werden verschiedene informatische Werkzeuge angewendet, um die gespeicherten Bioparameter mit den Parametern der definierten FHIR-Spezifikationen des Moduls „Intensivmedizin“ der Medizininformatik-Initiative (MII) zusammenzuführen. Am Ende dieser Zusammenführung entsteht ein Grundgerüst für die zukünftige Überführung der Biosignaldaten in FHIR.

Einleitung

Ein wichtiger Bestandteil eines sozialen Wandlungsprozesses ist der Prozess der Digitalisierung im Gesundheitswesen, denn die generierte Datenmenge ist heutzutage nicht im Papierformat zu bewältigen. Dieser Prozess bringt Herausforderungen mit sich, die zu bewältigen sind, sodass die Nutzung moderner Informationstechnik (IT)-Technologien und Standards im Gesundheitswesen in Bezug auf eine Verbesserung der Versorgung und Forschung im Gesundheitssystem ermöglicht wird. Hinzukommend übt die Digitalisierung auch einen Einfluss auf die Entwicklung der Interaktion zwischen unterschiedlichen an der Gesundheitsversorgung beteiligten Instanzen aus.

Eine der zentralen Herausforderungen im Prozess der Digitalisierung im Gesundheitswesen ist, zusammen mit dem immens generierten Datenvolumen, die mangelnde Interoperabilität vieler Systeme. Denn viele Unternehmen haben eigene Lösungen für einzelne Komponenten hergestellt, sodass die Interaktion von Systemen an einem Standort oder die Kommunikation zwischen verschiedenen Standorten in vielen Fällen impraktikabel ist. Eine weitere Problematik der mangelnden Interoperabilität ist die dazu mangelnde Nutzbarkeit der Daten für die Versorgung sowie für die Forschung, die auch die Krankenversorgung in näherer Zukunft fördern könnte. Aus diesem Grund soll innerhalb dieser Masterarbeit der Weg für die Erschließung von Biosignaldaten aus der Routineversorgung an der Universitätsmedizin Mainz primär für die wissenschaftliche Nutzung aufgezeigt werden.

1.1 Problemstellung

In den Krankenhäusern wird eine Unmenge an Daten aus der Routineversorgung gesammelt. Unter diesen Daten befinden sich die Biosignale. Diese sind die Ergebnisse von Messungen oder Beobachtungen. Diese Biosignaldaten werden entweder manuell (Körpergröße, Kopfumfang) oder automatisch durch am Netz verbundene Geräte (Beatmungsdruck, Blutflussindex) erfasst.

Die erfassten Parameter werden im Idealfall digital in komplexen Systemen gespeichert. Diese Informationen werden in den meisten Fällen über längere Zeiträume aufgehoben, weil sie entscheidend für die therapeutischen Behandlungen sind.

Mit der Zeit werden in einem Krankenhaus neue Geräte angeschafft, das Krankenhauspersonal wird erneuert und neue Techniken werden angewendet, sodass bei jeder Erneuerung auch Änderungen in dem System der Speicherung der Daten vorgenommen werden. Außerdem besitzt jedes Krankenhaus seine eigene Systemlandschaft zur Steuerung der Informationen und viele dieser Systeme benutzen keine Standardformate oder Codesysteme bei der Speicherung und Übermittlung der Daten, und somit sind die Daten in den Gesundheitseinrichtungen nicht interoperabel.

Dadurch, dass die Erfassung der Biosignaldaten in den meisten Fällen nicht den etablierten Standards entspricht, könnten solche Daten mit der Zeit unbrauchbar werden oder im schlimmsten Fall verloren gehen. Diese Situation erschwert die Nutzbarkeit der Informationen für die Forschung und Krankenversorgung - nicht nur an einem Standort, sondern auch deutschlandweit.

1.2 Ziel der Arbeit

Mit der vorher präsentierten Problematik ist das Ziel dieser Arbeit die ersten Schritte zu machen, zur Gewährleistung der Interoperabilität der, für längeren Zeitraum in einem PDMS an der Universitätsmedizin Mainz, gespeicherten Biosignaldaten, denn diese Information liegt weder in einem Standardformat vor noch beinhaltet es standardisierte gesundheitliche Codesysteme.

Um dieses Ziel zu erreichen wird die Abbildung der Biosignaldaten des PDMS in den FHIR-Spezifikationen des Erweiterungsmoduls „Intensivmedizin“ des Kerndatensatzes der MII erstellt. Die Methode wird an erster Stelle für Forschungszwecke eingesetzt und sollte die Grundsteine für die Umsetzung und Etablierung des Prozesses am Ende dieses Projekts legen.

Zusätzlich werden verschiedene IT-Werkzeuge und Techniken angewendet, um die Biosignalparameter für die spätere Überführung in das Standardformat FHIR bereitzustellen, sodass diese Daten in der Zukunft nicht nur lokal, sondern auch national und international datenschutzkonform benutzt werden können.

Hiermit wird, unter anderem, das Versorgungsniveau von Patienten und Patientinnen ausgewertet und zusammen mit anderen Projekten wird die frühzeitige Erkennung von gesundheitlichen Störungen geschaffen [1].

1.3 Aufbau der Arbeit

Zu Beginn dieser Arbeit werden zunächst Ansätze und bestimmte Grundlagen zu der zentralen Thematik erläutert. Das Konzept der Interoperabilität, insbesondere der Fokus auf dem Gesundheitswesen und die Entwicklung des Kerndatensatzes der MII mit verschiedenen Erweiterungsmodulen, wird präsentiert. An dieser Stelle wird sich die Arbeit mit dem Erweiterungsmodul Intensivmedizin befassen. Des Weiteren werden die digitalen Standards für das Gesundheitswesen Health Level Seven (HL7) und FHIR diskutiert. In einer weiteren Sektion werden drei wichtige Codesysteme erläutert. Zusätzlich wird das PDMS mit Fokus auf das COPRA-System präsentiert. Dazu folgen zwei Sektionen, eine über das Thema Data Warehouse (DWH), und die andere behandelt das Data Mapping. Am Ende des ersten Kapitels befasst sich eine Sektion mit dem Thema Pattern Matching und zwei Varianten dessen.

Im Kapitel „Methode“ wird die Abfolge der Durchführung dieses Projekts dargestellt. Das Kapitel beginnt mit einer kurzen Beschreibung des Prozessablaufes. Des Weiteren folgt die Schilderung der drei behandelnden Schritte des Data Mappings und die Entstehung eines Datensatzes mit zugeordneten Parametern. Anschließend wird die Analyse der Maßeinheiten des Datensatzes und die Rolle der Einheiten für die Programmierung der Transformationsregeln präsentiert. Am Ende wird die Vorbereitung der Daten für Überführung der Information aus COPRA in FHIR erläutert.

Das Kapitel „Realisierung“ umfasst die konkrete Durchführung der Methoden zusammen mit Ereignissen, die währenddessen entstanden sind. Anschließend wird die Entwicklung der Datenbank (DB) für die Zwischenspeicherung und Bearbeitung der Daten präsentiert. Zunächst werden in zwei Sektionen die Analyse der Eigenschaften der FHIR-Profile und die Bearbeitung der Parameter der FHIR-Profile erläutert. Zwei Sektionen befassen sich mit den Eigenschaften des COPRA-Systems. Danach folgt die Analyse der Parameter beider Systeme, diesmal in Zusammenhang mit der Überführung der Biosignaldaten aus COPRA in FHIR. Zusätzlich wird der Prozess der Zuordnung der FHIR-Profile mit den Konfigurationsvariablen gezeigt. Eine kurze Sektion präsentiert die Validierung der Zuordnung der FHIR-Profile mit den Konfigurationsvariablen. Die Durchführung der Analyse der Maßeinheiten und die Behandlung davon für die Programmierung der Transformationsregeln wird weiterhin diskutiert. Die zwei letzten Sektionen befassen sich mit der Bereitstellung der Daten für die Überführung der Biosignaldaten in FHIR.

Im Kapitel „Ergebnisse und Diskussion“ erfolgt eine kritische Bewertung der Projektergebnisse. Die drei Sektionen dieses Kapitels befassen sich

mit den Ergebnissen, Nebenergebnissen und Herausforderungen beider angewandten Systeme.

Das Kapitel „Fazit und Ausblick“ fasst diese Arbeit zusammen und wirft ein Blick auf die mögliche Entwicklung in der Zukunft.

Theoretische Grundlagen

Im folgenden Kapitel werden die Grundlagen zu den wichtigsten Konzepten, Fördermaßnahmen und Technologien näher betrachtet, die innerhalb dieser Arbeit im weiteren Verlauf angewendet werden. Zu Beginn werden das Konzept der Interoperabilität und die Erweiterung im Bereich des Gesundheitswesens präsentiert. Die Diskussion der Fördermaßnahme „MII“ mit seinem Kerndatensatz und das Erweiterungsmodul Intensivmedizin folgt im Anschluss. Die IT-Standards, wie HL7-FHIR und die dazu benutzten Datenaustauschformate, Extensible Markup Language (XML) und JavaScript Object Notation (JSON), werden nachfolgend erläutert. In einer weiteren Sektion wird sich die Arbeit mit den standardisierten Codesystemen SNOMED CT, LOINC und IEEE beschäftigen. Das PDMS und das in Deutschland etablierte Werkzeug COPRA werden präsentiert. Ein Überblick über das DWH-System wird im weiteren Verlauf der Arbeit verschafft. Anschließend werden Aspekte des Data Mappings und Pattern Matchings in zwei Sektionen erläutert.

2.1 Interoperabilität

Der Begriff Interoperabilität kann je nach Anwendungsgebiet unterschiedliche Bedeutungen tragen. Die jedoch meist genutzte Definition ist die Folgende: „Fähigkeit von zwei oder mehreren Systemen oder Komponenten, Informationen auszutauschen und die ausgetauschten Informationen wieder zu nutzen.“ [2]

Diese Definition beinhaltet zwei Ebenen der Interoperabilität, nämlich den Austausch von Information (technische Interoperabilität) und die Nutzung der ausgetauschten Information (semantische Interoperabilität) [3]. Die dritte Ebene ist die Prozessinteroperabilität [4]. Dies wird erreicht, wenn Menschen gemeinsames Verständnis über ein Netzwerk teilen, Systeme zusammenarbeiten und Arbeitsabläufe koordiniert werden [5]. Im Gesundheits-

wesen wird auch von der klinischen Interoperabilität gesprochen [4]. Diese Ebene ist die Fähigkeit mehrere klinische Fachkräfte in unterschiedlichen Versorgungsteams einzusetzen, um eine nahtlose Versorgung von Patienten und Patientinnen zu gewährleisten [5].

2.1.1 Interoperabilität im Gesundheitswesen

Seit einigen Jahren eröffnet die steigende Vernetzung zusammen mit der Digitalisierung neue Möglichkeiten in der Patientenbetreuung [3]. Denn es ermöglicht eine neue Art der Kommunikation zwischen dem Personal des Gesundheitssystems an verschiedenen Standorten als auch die Kommunikation mit Patienten und Patientinnen [4]. Solche Kommunikation benötigt eine organisatorische und inhaltliche Harmonisierung der zur Verfügung gestellten Daten in Konkordanz mit den neuesten IT-Standards und Schnittstellen, um die Interoperabilität zu gewährleisten [3].

Ein wichtiger Aspekt auf dem Weg der Digitalisierung, und somit der Verbesserung des Gesundheitssystems, sind verbindliche Klassifikationen und Ontologien für eine eindeutige und fachliche Kommunikation [4]. Diese interoperablen Mittel sollen beim Austausch von Daten und die Implementierung von IT-Lösungen umgesetzt werden. Um dieses Ziel zu erreichen, wurde in Deutschland im Jahr 2016 vom Bundesministerium für Bildung und Forschung (BMBF) die MII gegründet [3].

Die Aufgabe der MII ist die Verbesserung von Forschungsmöglichkeiten und Patientenversorgung durch IT-Lösungen [6]. Damit werden Forschung und Versorgung innerhalb und zwischen den Universitätskliniken, Forschungseinrichtungen, Unternehmen, Krankenkassen und Patientenvertreter vernetzt [3, 6]. In dieser Vernetzung spielen Schnittstellen medizinischer Inhalte mit HL7 (2.3) eine sehr wichtige Rolle für den Austausch der medizinischen Daten [3].

Die Mitglieder der MII haben sich auf einen Kerndatensatz (2.2) geeignet, der auf Basis internationaler Standards, wie SNOMED CT (2.4.1), LOINC (2.4.2), etc. entwickelt wurde, um die Interoperabilität zwischen den Standorten zu garantieren [3, 7]. Dieser Datensatz beinhaltet zusätzlich verschiedene Erweiterungsmodule für unterschiedliche Use Cases [6]. Diese Spezifikation ist verbindlich für die syntaktische und semantische Kodierung des Inhaltes der Module [1].

Das Kernelement der MII sind die Datenintegrationszentren (DIZs), dessen Herausforderung die Aufnahme, Zusammenführung und Aufbereitung der Daten aus verschiedenen Systemen, sowie die Sicherstellung von Datenqualität und Datenschutz dieser Daten ist [6, 8]. Damit können die Daten in Versorgung und Forschung genutzt werden. Um dieses Ziel zu erreichen, wer-

den die Daten standardisiert, wiederverwendbar und austauschbar gemacht [8]. Jedes Mitglied der MII besitzt ein lokales DIZ, in dem die lokalen Daten gespeichert werden. Bei verteilten Machbarkeitsabfragen, wie z. B. die Übermittlung der Menge an behandelnden Personen mit einer bestimmten Diagnose, wird die gespeicherte Information datenschutzkonform bereitgestellt, und in aggregierter Form dann zentral bewertet.

2.2 Kerndatensatz der MII

Einer der Erfolge der MII ist, dass alle Mitglieder der MII sich auf einen gemeinsamen Kerndatensatz geeinigt haben [3, 7]. Dieser Datensatz ist die wichtigste Voraussetzung für die zentrale Basis für den gemeinsamen Gebrauch von Daten und, dass dieselbe Auswertungslogik in den DIZs lokal ausgeführt werden kann, denn die Daten in der Patientenversorgung stammen heutzutage aus diversen Quellsystemen und liegen somit in vielen verschiedenen Datenformaten mit unterschiedlichem Inhalt vor [7].

Mit diesem Kerndatensatz wurde, unabhängig von Use Case und Indikation der Mitglieder der MII, festgelegt, welche Datensätze von den stationären und ambulanten Patienten und Patientinnen die DIZs der MII mindestens speichern sollten [7].

Der MII-Kerndatensatz besteht aus Basis- und Erweiterungsmodulen (Abbildung 2.1). Während die Definition der Basismodule fachlich übergreifend ist, sind die Erweiterungsmodule die Abbildung von Daten spezifischer Anwendungs- oder Fachgebiete, sodass der Kerndatensatz in kontinuierlicher Erweiterung ist [7].



Abbildung 2.1: Kerndatensatz der MII und deren Erweiterungsmodule [7].

2.2.1 Erweiterungsmodul „Intensivmedizin“ des Kerndatensatzes der MII

Relevant für die Durchführung dieser Masterarbeit ist das Erweiterungsmodul „Intensivmedizin“ oder Intensive Care Unit (ICU), wie in Abbildung 2.1 als PDMS/Biosignale genannt. Der Fokus des Projekts liegt ganz konkret auf der FHIR-Spezifikation des Moduls (Medizininformatik Initiative - Modul ICU - ImplementationGuide). Ein wichtiger Hinweis an dieser Stelle ist, dass das Modul „Intensivmedizin“ auch eine parallele Webseite für die Weiterentwicklung unter der Open Source Plattform SIMPLIFIER.NET (Medizininformatik Initiative - Modul ICU) besitzt, wo die aktuellen Modifikationen veröffentlicht werden. Auf dieser Webseite können die Mitglieder der MII und weitere Personen mit Zugang zu dieser Webseite Kommentare und Anmerkungen an die SIMPLIFIER-Webseite des Moduls (Issues) senden.

Dieses Erweiterungsmodul spezifiziert akutmedizinische Daten für die Primär- und Sekundärnutzung und hat Bezüge zu den Basismodulen (Abbildung 2.1). Die erste stabile Ballot-Version der FHIR-Profile des Moduls wurde am 24. Februar 2022 veröffentlicht [9]. Ziel der Modellierung dieses Erweiterungsmoduls ist die Datenabbildung der Intensivmedizin und die Darstellung gleichartiger Daten der Notfallmedizin, stationärer und ambulanter Medizin [1]. Ziel des Moduls ist die Unterstützung und Erleichterung der Forschung, Qualitätssicherung, Kommunikation zwischen medizinischen Geräten

und der Entscheidungsfindung [10].

Das Erweiterungsmodul „Intensivmedizin“ ist in drei Entwicklungsstufen geplant, davon ist die erste Stufe fertiggestellt, die zweite wurde von 25.02.2022 bis 08.04.2022 zum Abstimmverfahren freigegeben¹, es ist ein Entwurf zur Korrektur und zum Kommentieren; die dritte Stufe befindet sich immer noch in Entwicklung [10].

1. Entwicklungsstufe 1: Datenmodell

2. Datenarten der Entwicklungsstufe 2:

- Monitoring und Vitaldaten
- Parameter von extrakorporalen Verfahren
- Beatmungswerte

3. Datenarten der Entwicklungsstufe 3:

- Ein-/Ausfuhrbilanzen
- Scores
- Hochauflösende Daten
- Lagerungstherapie
- etc.

Der Schwerpunkt dieses Projekts liegt in den Parametern der Entwicklungsstufe 2 des Moduls. Obwohl der Abstimmungsprozess dieser Stufe abgeschlossen ist, wurde bis jetzt noch keine neue überarbeitete Version der Entwicklungsstufe 2 fertiggestellt.

Die Datenarten der Stufe 2 besitzen mindestens ein generisches FHIR-Profil und viele davon auch ein spezifisches Profil, welches wenn möglich verwendet werden sollte [1, 10]. Die generischen Profile modellieren die Daten unabhängig von einer spezifischen Annotation und beschreiben die Struktur für Gruppen von Elementen einer bestimmten Kategorie und sind somit die ersten in der Baumstruktur der FHIR-Profile (Abbildung 2.2) [1]. Die spezifischen FHIR-Profile hingegen modellieren zusätzlich eine eindeutige semantische Annotation [10].

Unter der Gruppe der Parameter von extrakorporalen Verfahren befinden sich die erfassten Daten von Extracorporeal Membrane Oxigenation

¹Am 07.11.2022 wurde die finale Version des Moduls freigegeben. Dieses Projekt wurde aber mit der ersten stabilen Ballot-Version des Moduls realisiert, welche in Abstimmung war.

(ECMO)-Maschinen, wie die Dauer einer Hämodialysesitzung. In dieser Gruppe werden die gemessenen und eingestellten Werte zusammen mit dem ECMO-Gerät erfasst [1].

Die Beatmungswerte der Stufe 2 umfassen Daten, wie die Sauerstofffraktion, das Beatmungsvolumen, etc. Der Aufbau dieser FHIR-Profile ist analog zu Parametern von extrakorporalen Verfahren [1].

Die FHIR-Profile unter der Gruppe Monitoring und Vitaldaten umfassen Daten, wie die Herzfrequenz, Körpergröße, die pulsatilen Blutdrücke (systolisch, mittel und diastolisch), etc. Für die pulsatilen Blutdrücke wurde ein weiteres generisches Profil (Blutdruck Generisch) erstellt, um die Kompatibilität mit den deutschen und internationalen Basisprofilen zu gewährleisten [1]. Diese Profile sind auch eine Zusammensetzung von drei separat kodierten „Observations“, die dieselben Attribute teilen, nämlich die Spezifikationen für den systolischen, mittleren und diastolischen Blutdruck [1].



Abbildung 2.2: Baumstruktur des Erweiterungsmoduls „Intensivmedizin“.

Die semantische Annotation des Erweiterungsmoduls „Intensivmedizin“ referenziert mindestens einen Primärcode der Terminologie (SNOMED CT oder LOINC) [1, 10]. Um die Interoperabilität mit der Kommunikation zwischen Medizingeräten oder Medizinprodukten zu ermöglichen, wird wenn möglich auch eine semantische Annotation nach ISO/IEEE 11073-10101TM (2.4.3) hinzugefügt [1]. Die Maßeinheiten werden nach Unified Code for Units of Measure (UCUM) codiert.

Die FHIR-Profile des Moduls ICU sind in drei Kategorien untergliedert:

- DeviceMetric
- Observation
- Procedure

Die FHIR-Profile der Kategorie „DeviceMetric“ beschreiben die Messungen und die Einstellungen des benutzten medizinischen Geräts, und sollen wenn möglich mit dem benutzten Gerät verlinkt werden [1]. „DeviceMetric“-Profile kategorisieren die Werte in gemessenen und eingestellten Werten [11]. Die gemessenen Werte sind die während der Untersuchung erhobene Parameter, z. B. die spontane mechanische Atemfrequenz bei der Beatmung. Die eingestellten Werte sind wiederum die Einstellungen an den Geräten, z. B. der eingestellte inspiratorische Gasfluss während der Beatmung.

Die FHIR-Profile unter der Kategorie „Observation“ registrieren die Messungen an Patienten, wie Laboruntersuchungen oder die Beobachtungen an Geräten [12]. Unter dieser Kategorie befinden sich die Vitalparameter.

„Procedure“-Profile definieren ein durchgeführtes Verfahren an Patienten, diese können entweder eine physikalische Intervention oder eine Beratung sein [13].

2.3 HL7 & FHIR

Wichtig um die Interoperabilität in Gesundheitswesen zu gewährleisten, sind syntaktische IT-Standards wie FHIR, der neueste Standard von HL7 mit Fokus auf den aktuellen Web-Standards und einer einfachen Implementierung [3, 14, 15].

2.3.1 HL7

Health Level Seven (HL7) ist eine nicht gewinnorientierte, American National Standards Institute (ANSI)-akkreditierte Standardisierungsorganisation zur Bereitstellung eines umfassenden Frameworks und dessen Standards auf der 7. Ebene (Anwendungsschicht) im ISO/Open Systems Interconnection (OSI) Modell, für den Austausch, die Integration, das Teilen und den Abruf elektronischer Gesundheitsinformationen [3, 4, 14].

HL7-Schnittstellen unterstützen die Kommunikation von Softwaresystemen in klinischen Einrichtungen und somit das Management, die Erbringung und Bewertung von Gesundheitsdiensten [4, 14, 15].

2.3.2 FHIR

Der Fast Healthcare Interoperability Resources (FHIR)-Standard eignet sich für den Einsatz in verschiedenen Szenarien, wie den Datenaustausch auf nationaler und internationaler Ebene, in einem regionalen Netzwerk, zwischen Systemen innerhalb einer Organisation und den Datenaustausch mit mobilen Applikationen [4, 15].

FHIR besteht aus drei Bausteinen: Ressourcen (Resources), Referenzen (References) und Profile (Profiles) [15].

In FHIR werden Konzepte aus der realen Welt als Ressourcen mit entsprechenden für Menschen und Maschinen lesbaren Werten dargestellt [4]. Diese wiederum sind logische Einheiten des Datenaustausches mit einem konkreten, definierten Verhalten und eindeutiger Semantik [15]. Die Definition der Ressourcen deckt häufig die meisten Anwendungsfälle [4]. Jede Ressource besteht aus strukturierten Daten, Narrativen und Ergänzungen (Extensions) [15]. Während die strukturierten Daten 80% der weitverbreiteten Einsatzszenarien abdecken, lassen sich manche Ressourcen durch die Ergänzungen erweitern, die nicht in FHIR erhältlich sind, um speziellere Anwendungsfälle abdecken zu können [15, 4]. Die Narrative sind dann menschenlesbare Texte, um den Inhalt der Ressourcen zusammenzufassen [15].

Die Referenzen ermöglichen es, dass mehrere Ressourcen und deren Informationseinheiten aufeinander verweisen können, um Beziehungen aufzubauen [15]. Ein Beispiel dafür sind die „Observation“-Ressourcen im Erweiterungsmodul „Intensivmedizin“. Diese verweisen auf die „Patient“-Ressourcen im Basismodul „Person“ des Kerndatensatzes der MII (Abbildung 2.1).

Viele Werte in den FHIR-Ressourcen sind optional, sodass eine Umsetzung kompatibel ist, wenn sie nur die minimal obligatorischen Daten liefert. Diese Werte können von außen über die Profile modifiziert werden, und können wiederum sich aufeinander verweisen [4]. Das bedeutet die FHIR-Profiles sind die Spezifikationen, die die FHIR-Ressourcen beschreiben und definieren, z. B. die Festlegung von konkreten Codesystemen [15, 16].

Ein zuvor genannter Aspekt von FHIR ist seine Fokussierung auf moderne Webprotokolle. Somit können die FHIR-Ressourcen, die auf einem Server liegen, über eine Schnittstelle via Hypertext Transfer Protocol (HTTP) abgefragt werden [3, 4]. Außerdem ist das Datenformat für diese Übertragung auch von FHIR festgelegt, diese kann entweder XML oder JSON sein [4]. Diese zwei Formate sind weitverbreitet und besitzen eine hohe Akzeptanz durch ihre Versatilität [17].

Die FHIR-Profiles können auf verschiedene Art und Weise dargestellt werden. Die am häufigsten benutzten Arten der Darstellung sind logische Tabellen, XML- und JSON-Schemata [17].

Logische Tabelle

Eine logische Tabelle eines FHIR-Profiles beinhaltet die Spalten: Name, Flags, Card., Type und Description & Constraints, wobei diese letzte Spalte beim Anklicken eines Elements eingeblendet wird [17]. Abbildung 2.3 zeigt ein Teil der logischen Tabelle eines FHIR-Profiles des Erweiterungsmoduls „Intensivmedizin“.

Name	Flags	Card.	Type	Description & Constraints
Procedure	I		Procedure	
extension	I	0..*	Extension	
identifier	Σ	0..*	Identifier	
instantiatesCanonical	Σ	0..*	canonical(PlanDefinition ActivityDefinition Me...	
instantiatesUri	Σ	0..*	uri	
basedOn	Σ I	0..*	Reference(CarePlan ServiceRequest)	
partOf	Σ I	0..*	MII-Reference(Procedure Observation)	
status	S Σ ?!	1..1	code Binding	
statusReason	Σ	0..1	CodeableConcept	
category	S Σ	1..1	CodeableConcept	
code	S Σ I	1..1	CodeableConcept	
subject	Σ I	1..1	MII-Reference(Patient)	
encounter	S Σ I	0..1	MII-Reference(Encounter)	
performed[x]	S Σ	1..1		

There are no (further) constraints on this element

Element Id

Procedure.code

Short description
Identification of the procedure

Alternate names
type

Definition
The specific procedure that is performed. Use text if the exact nature of the procedure cannot be coded (e.g. "Laparoscopic Appendectomy").

Requirements

Abbildung 2.3: Fragment der logischen Tabelle des FHIR-Profiles „Beatmung“ des Erweiterungsmoduls „Intensivmedizin“ des Kerndatensatzes der MII.

- Name: Name eines Elements in der FHIR-Ressource. Das Icon bezeichnet den Inhalt des Elements. In der Abbildung 2.4 wird die Bedeutung der Icons erklärt
- Flags: Entscheidende Informationen darüber, wie das Element implementiert werden sollte
 - S: Dieses Element muss unterstützt werden
 - I: Elemente mit diesem Symbol definieren Randbedingungen oder sind Teile davon, z. B. das Element `code` definiert die benutzten Codesysteme bei der „Procedure“-Profile (Abbildung 2.3)
 - Σ: Diese Elemente sind Teile einer Ressource
 - ?!: Das sind modifizierende Elemente, z. B. das Element `status` beschreibt den Zustand einer „Procedure“-Ressource, nämlich `final`, wenn die Ressource in einer Endversion sich befindet (Abbildung 2.3)

- Card.: Cardinality - untere und obere Grenze, wie häufig dieses Element in der Ressource erscheinen darf
- Type: Typ des Elements. Hier erscheint auch der Link zur Definition des Typs
- Description & Constraints: Beschreibung des Elements und Details über die Einschränkungen

-  : Das Basiselement für eine Ressource
-  : Ein Element in der Ressource, das in der Ressource oder dem Profil definiert ist
-  : Ein Element, das einen von mehreren unterschiedlichen Typen haben kann
-  : Ein Element eines Datentyps, der ein Element mit Wert oder Eigenschaft beschreibt
-  : Element eines Datentyps, der ein Element mit in sich tragendem Element beschreibt
-  : Eine Extension, die einen Wert und keine verschachtelten Extensions hat
-  : Ein Element mit einer Referenz zu einer anderen Ressource

Abbildung 2.4: Beschreibung der meist benutzten FHIR-Icons.

XML

Die Metasprache Extensible Markup Language (XML) hat ihre Wurzeln im Internet und wurde vom World Wide Web Consortium (W3C) im Jahr 1998 veröffentlicht und dessen aktuelle Fassung ist die fünfte von 2008 [18]. Als Metasprache können mit Hilfe von XML andere anwendungsspezifische Sprachen, wie FHIR, definiert werden. Mit XML können auch Daten gespeichert oder übertragen werden [5, 18].

Ein XML-Dokument ist für das Computer-Processing konzipiert, trotzdem kann dieses Dokument von Menschen gelesen werden, denn Dokumente im XML-Format sind zur selben Zeit Text-Dateien [18]. Die Regeln bei XML-Dateien sind strikt, sodass Fehler in den Anwendungen verhindert werden. Alle XML-Dokumente beinhalten verbundene Knoten in einer Baumstruktur, wo alle Knoten von einer einzelnen Wurzel stammen, das so genannte Element [5]. Für die Auszeichnung von Elementen werden Tags „<, >, />“ verwendet [18]. Attribute werden in XML benutzt, um die Information der Elemente einzufügen [5, 19]. Der Code 2.1 zeigt die FHIR-XML-Datei für die Definition einer Person.

```

<?xml version="1.0" encoding="UTF-8"?>

<Patient xmlns="http://hl7.org/fhir">
  <id value="xds"/>
  <text>
    <status value="generated"/>
  </text>
  <identifier>
    <use value="usual"/>
    <type>
      <coding>
        <system value="http://terminology.hl7.org/CodeSystem/v2-0203"/>
        <code value="MR"/>
      </coding>
    </type>
    <system value="urn:oid:1.2.3.4.5"/>
    <value value="89765a87b"/>
  </identifier>
  <active value="true"/>
  <gender value="male"/>
  <birthDate value="1989-05-27"/>
  <address>
    <postalCode value="55122"/>
  </address>
  <managingOrganization>
    <reference value="Organization/2"/>
  </managingOrganization>
</Patient>

```

Listing 2.1: Beispiel einer FHIR-Ressource für eine Person in XML.

Die grundlegende analytische und Design-Aufgabe der Nutzung von XML ist die Erstellung von Schemata [18]. Die Struktur eines XML-Dokuments wird in einem Schema beschrieben, das wiederum in XML geschrieben ist. Das Schema bestimmt die Struktur eines Dokumenttyps, die für alle Dokumente dieser Art gleich ist [5, 18]. Es legt die Tags und die Verbindungen zwischen den Elementen fest. Die Schemata werden anhand eines oder mehrerer Schemata mit Hilfe von Anwendungen validiert [18]. In HL7 sind zwei Schemasprachen häufig verbreitet, die Schema Definition Language (XSD): XML-Schemata des W3C zur Strukturdefinition von XML-Dokumenten, und Schematron zur Validierung von Inhalt und Struktur von XML-Dokumenten [5]. XML-Schemata werden normalerweise als separate Datei mit der Erweiterung .xsd erstellt und über eine Namespace-Deklaration mit der Datei verlinkt [5].

JSON

JavaScript Object Notation (JSON) ist ein, auf Features der Skriptsprache JavaScript basiertes, menschlesbares Datenaustauschformat für die Übertragung der Information zwischen Anwendungen [20]. Aus diesem Grund kann die Sprache selbst die Daten in JSON-Format decodieren, und die Daten können als native JavaScript-Objekte direkt angewendet werden [5].

Dadurch, dass JSON einfacher als XML ist und viele Programmiersprachen schon JSON-Bibliotheken implementiert haben, ist JSON das bevorzugte Datenaustauschformat zwischen Anwendungen im unterschiedlichen Kontext, z. B. FHIR-Server und Webanwendungen, und damit eine geeignete Option, um Interoperabilität in Gesundheitswesen zu gewährleisten [5, 20]. Der Code 2.2 zeigt die FHIR-JSON-Datei für die Definition einer Person.

```
{
  "resourceType": "Patient",
  "id": "xds",
  "text": {
    "status": "generated",
  },
  "identifier": [
    {
      "use": "usual",
      "type": {
        "coding": [
          {
            "system": "http://terminology.hl7.org/CodeSystem/v2-0203",
            "code": "MR"
          }
        ]
      },
      "system": "urn:oid:1.2.3.4.5",
      "value": "89765a87b"
    }
  ],
  "active": true,
  "gender": "male",
  "birthDate": "1989-05-27",
  "address": [
    {
      "postalCode": "55122",
    }
  ],
  "managingOrganization":
```

```

{
  "reference": "Organization/2"
}

```

Listing 2.2: Beispiel einer FHIR-Ressource für eine Person in JSON.

JSON besteht aus sechs Datentypen für die Darstellung von Entitäten der realen Welt [20]:

1. String: Unicode-Zeichen
2. Number: Gleitkomma-Format von JavaScript mit doppelter Genauigkeit
3. Boolean: Die Werte `true` und `false` ohne Anführungszeichen
4. Null: Leerer Wert
5. Object: Gruppe von Namens- oder Wertpaaren zwischen `{ }`
6. Array: Geordnete Sammlung von Werten

2.4 Codesysteme

Um die semantische Interoperabilität im Gesundheitswesen zu gewährleisten, werden auch Codesysteme benötigt. Ein Codesystem ist eine organisierte Sammlung von Begriffen, in der jeder Begriff durch mindestens einen intern eindeutigen Code dargestellt wird [21]. Weiterhin kann ein Codesystem auch eine sprachabhängige Definition enthalten, wobei einige Konzepte sehr spezifisch sind, andere hingegen können sehr allgemein sein [5, 21].

Codesysteme sind ein wesentlicher Teil bei den medizinischen Anwendungen und den interoperablen Spezifikationen für den Austausch von Daten zwischen Computern [5]. Außerdem definieren die Codesysteme, welche Symbole oder Ausdrücke existieren und wie sie zu verstehen sind [5, 22].

Manche Codesysteme beinhalten komplexe Ideen mit vielen Subklassifikationen, wie die International Statistical Classification of Diseases and Related Health Problems (ICD), andere wiederum sind weit gefasst, wie SNOMED CT (2.4.1), oder sind auf spezifische Bereiche fokussiert, wie LOINC (2.4.2), weiterhin überlappen sich diverse Codesysteme miteinander, wie ICD und SNOMED CT [21].

2.4.1 SNOMED CT

Weltweite Standards für Gesundheitsbegriffe werden von der International Healthcare Terminology Standards Development Organization (IHTSDO) bestimmt, weiterhin ist die IHTSDO für die Pflege, Weiterentwicklung, Qualitätssicherung und Veröffentlichung des ontologiebasierten Terminologiestandards SNOMED CT verantwortlich [23].

Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) ist die umfangreichste mehrsprachige klinische Gesundheitsterminologie, so dass sie in elektronischen Patientenakten verwendet wird [3, 5]. Darüber hinaus erleichtert die Nutzung von SNOMED CT die klinische Dokumentierung und die Berichterstattung [3]. Außerdem hilft SNOMED CT bei der Ermittlung und Analyse von klinischen Daten und darüber hinaus ist SNOMED CT sowohl ein Codierschema zur Ermittlung von Begriffen und Konzepten, als auch eine mehrdimensionale Klassifizierung, die es ermöglicht, Konzepte miteinander in Verbindung zu bringen, zu gliedern und nach verschiedenen Kategorien zu analysieren [5].

Das logische Datenmodell von SNOMED CT (Abbildung 2.5) beinhaltet drei Komponenten: Konzepte (Concepts), Beschreibungen (Descriptions) und Relationen (Relationships), die sich durch Referenz Sets (RefSets) ergänzen, außerdem sind alle SNOMED CT-Komponente durch eine eindeutige numerische Kennung identifizierbar [24].

- Die Konzepte repräsentieren medizinische Begriffe und werden hierarchisch von allgemein zu speziell angeordnet.
- Die Beschreibungen verknüpfen medizinische Begriffe mit Konzepten. Außerdem beinhaltet ein Konzept eine eindeutige, unmissverständliche Beschreibung dessen Bedeutung und kann auch weitere Beschreibungen erhalten (Synonyme). Dazu enthält jede Übersetzung der SNOMED CT eine zusätzliche Menge von Beschreibungen, die Fachbegriffe in anderen Sprachen mit demselben Konzept verlinken.
- Die Relationen verknüpfen zwei Konzepte miteinander. Ein Relationstyp oder Attribut wird benutzt, um die Bedeutung der Verbindung zwischen Ausgangskonzept und Zielkonzept zu repräsentieren, weiterhin verbindet der Relationstyp „is-a“ ein Konzept zu allgemeineren Konzepten und bildet so die Konzepthierarchien in SNOMED CT, andere Relationstypen hingegen repräsentieren Aspekte eines Konzepts.
- Die RefSets sind standardisierte SNOMED CT-Implementierungen für die Unterstützung von Anforderungen für Anpassungen und Erweiterungen.



Abbildung 2.5: SNOMED CT-Datenmodell.

Die drei SNOMED CT-Komponenten des Elements Blutdruck (Blood pressure) werden in der Tabelle 2.1 repräsentiert.

Tabelle 2.1: Blutdruck in SNOMED CT.

Konzept	Beschreibung	Relation
ID: 75367002 - Blood pressure	Blood pressure (observable entity)	Is a → ID: 310611001 - Cardiovascular measure

2.4.2 LOINC

Das Logical Observation Identifiers Names and Codes (LOINC)-System ist seit 1994 eine internationale Zusammenstellung von Begriffen (Terminologie) zur eindeutigen Identifizierung und Kodierung für Laboruntersuchungen, klinische und medizinisch-technische Untersuchungen, medizinische Dokumententypen und Fragebögen/Fragen [25, 26]. Somit ermöglicht LOINC das Zusammenführen von Untersuchungsergebnissen durch standardisierte Bezeichnungen für den Datenaustausch im Gesundheitswesen und die elektronische Kommunikation von Daten zwischen Labor, Klinik und Praxis [5]. LOINC erstellt Codes und einen formalen Namen für jedes Konzept, das einer einzelnen Art von Beobachtungsmessung oder Testergebnis entspricht [5]. Dieser formale LOINC-Name ist vollständig spezifiziert, sodass er die Merkmale zur erforderlichen Unterscheidung zwischen ähnlichen und unterschiedlichen klinischen Beobachtungen enthält [3]. HL7-FHIR bietet auch Unterstützung für LOINC [26].

Die LOINC-DB stellt einen Satz von eindeutigen Namen und Identifikatoren für die Identifikation von Laboruntersuchungen und weiteren klinischen Testen zur Verfügung [27]. In dieser DB befinden sich zusammen mit den LOINC-Coden weitere Parameter, die eine Untersuchung beschreiben, z. B. Maßeinheiten, „Short Name“ oder Alias und Basis Attribute, wie der Typ der Untersuchung.

Die Struktur eines LOINC-Codes beinhaltet heutzutage zwischen drei und sieben Zeichen in der Form „Ziffern-Ziffer“. Das „-“ befindet sich immer an der vorletzten Stelle und die letzte Ziffer des Codes befindet sich zwischen 0 und 9. Diese letzte Zahl ist eine Prüfzahl für den LOINC-Server, um Fehler bei der Übertragung des Codes zu vermeiden. Diese Menge an Zeichen kann mit dem Wachstum der LOINC-DB vergrößert werden [28].

Ein LOINC-Code ist durch sechs Dimensionen in der folgenden Reihenfolge definiert [25]:

- Component oder Analyte: Gemessene oder beobachtete Substanz oder Entität
- Property: Merkmal oder Attribut der Substanz oder Entität
- Time: Zeitintervall, in dem eine Beobachtung gemacht wurde
- System: Beobachtetes System (Probenmaterial, Körperteil oder Umgebung)
- Scale: Wie der Beobachtungswert quantifiziert oder ausgedrückt wird: quantitativ, ordinal oder nominal
- Method: Methode, mit der die Messung oder Beobachtung stattgefunden hat

Die Dimension „Method“ ist operationell und wird nur verwendet, wenn die Technik die klinische Interpretation der Ergebnisse beeinflusst [5]. Ein Beispiel der LOINC-Dimensionen ist in der Tabelle 2.2 dargestellt.

Tabelle 2.2: LOINC-Dimensionen des pulmonalvaskulären Widerstandsindex. LOINC-Code: 8834-4. ArResis steht für Resistance/Area (Widerstand/Fläche), Pt für Point time (Zeitpunkt-Messungen) und Qn für Quantity. Diese Entität besitzt keine Methode, die die Interpretation der Ergebnisse beeinflusst.

Component	Property	Time	System	Scale	Method
Hemodynamic resistance/Body surface area	ArResis	Pt	Pulmonary vasculature	Qn	-

2.4.3 ISO/IEEE 11073™

Das Institute of Electrical and Electronics Engineers (IEEE) ist die weltweit größte technische Fachorganisation, die sich für den Technologiefortschritt zum Nutzen der Bevölkerung einsetzt, und entwickelt somit International Organization for Standardization (ISO) Normen, die die Grundlage für viele der heutigen Produkte und Dienstleistungen in der Informationstechnologie darstellen [29]. Die verabschiedeten Standards des IEEE sind häufig die zentrale Quelle für die Standardisierung eines breiten Spektrums aufstrebender Technologien [30]. Eine dieser Familien von Standards ist die ISO/IEEE 11073™ [30, 31]. Diese Serie definiert die Komponente eines Systems, die den Austausch und die Auswertung von Vitaldaten zwischen medizinischen Geräten, zusammen mit der Fernsteuerung solcher Geräte, ermöglichen [31]. Weiterhin sind diese Standards ein Ansatz, um das Problem der nicht interoperablen Medizinprodukte zu lösen.

ISO/IEEE 11073-10101™ Nomenklatur

Die ISO/IEEE 11073-10101™ Nomenklatur wurde von dem ISO/IEEE 11073™ Standard Komitee der „IEEE Engineering in Medicine and Biology Society“ vorbereitet [29]. Die ISO/IEEE 11073-10101™ Nomenklatur legt Nomenklaturcodes fest, die die Interoperabilität von Medizinprodukten auf semantischer Ebene erleichtern [31]. Dieser Standard ist auf medizinischen Geräten für die Akutversorgung und Informationen über die Vitalparameter der Patienten und Patientinnen fokussiert [29]. Die ISO/IEEE 11073-10101™ definiert sowohl die Architektur als auch die Hauptkomponenten der Nomenklatur, zusammen mit ausführlichen Definitionen für jeden konzeptionellen Bereich [32]. Dieser Standard ermöglicht die Verlinkung mit LOINC und SNOMED CT, außerdem werden auch FHIR-Extensions entwickelt, die ISO/IEEE 11073-10101™ unterstützen [33]. Der Abschnitt von JSON-Code 2.3 zeigt die Anwendung von einem ISO/IEEE 11073-10101™-Identifikator im FHIR-Profil „Atemzugvolumen-Während-Beatmung“ des Erweiterungsmoduls „ICU“ des Kerndatensatzes der MII.

```
{
  "system": "urn:iso:std:iso:11073:10101",
  "code": "151980"
}
```

Listing 2.3: Beispiel der Anwendung von ISO/IEEE 11073-10101™ in FHIR.

2.4.4 UCUM

Der UCUM ist ein sehr stabiles Codesystem, das alle gegenwärtigen in der internationalen Wissenschaft, Technik und Wirtschaft verwendeten Maßeinheiten umfassen soll [34]. Das Ziel des UCUM besteht darin, die eindeutige elektronische Übermittlung und Interpretation von Größen zusammen mit ihren Maßeinheiten zu erleichtern [34, 35]. Aus diesem Grund wurde das Codesystem UCUM bereits von Standardorganisationen wie HL7 übernommen [35].

2.5 Patientendatenmanagementsystem

Das Patientendatenmanagementsystem (PDMS) unterstützt die klinische Dokumentation auf Intensivstationen „Intensive Care Unit (ICU)“ und hat damit nachweisbare Auswirkungen auf die Vollständigkeit der Patientenakten, den Zeitaufwand für die Dokumentation und die Erhöhung der Patientenqualitätssicherung [36, 37, 38]. Dieses System umfasst auch Komponenten der computergestützten Auftragserfassung für Ärzte „Computerized Physician Order Entry (CPOE)“, denn die Daten aus diesem System sind entscheidend für verschiedene automatisierte Workflows [36, 38]. Das PDMS bietet auch eine spezifische Funktionalität für die Dokumentation der Diagnosis Related Group (DRG), welche alle relevanten Daten für die Codierung erfasst [36, 37].

In Deutschland waren im Jahr 2021 ca. 15 PDMS-Anbieter bekannt [39]. Unter diesen Anbietern befindet sich der Computer Organized Patient Report Assistant (COPRA), welcher an der Universitätsmedizin Mainz und Standort der Realisierung dieses Projekts benutzte PDMS ist.

2.5.1 COPRA

COPRA System GmbH ist seit 1993 einer der führenden Anbieter von PDMSs in Deutschland [40, 41]. Dessen Hauptprodukt ist das zertifizierte Medizinprodukt COPRA in der Version 6 „COPRA6“. Mit seinen vier Anwendungsgebieten: Ärzte und Ärztinnen, Pflege, Controlling und IT-Abteilung ist COPRA ein PDMS für die Dokumentation von Behandlung und Pflege geeignet [41]. Aus diesem Grund wird COPRA als PDMS seit 2007 an der Universitätsmedizin Mainz etabliert [42].

Für die Ärzte und Ärztinnen bietet das COPRA-System eine transparente und vereinfachte Möglichkeit für die Dokumentation aller relevanten Daten einer Behandlung, denn alle Befunde der behandelnden Personen können in

COPRA eingesehen werden, darüber hinaus werden ärztliche Anordnungen dokumentiert und freigeschaltet [41].

Mit den sogenannten Arbeitslisten für die Pflege werden die bereits erfolgten und nicht erfolgten Behandlungsschritte in COPRA angezeigt. Außerdem werden die Kurven der Vitalparameter durch die automatische Übernahme von Werten aller an Patienten angeschlossenen Geräte aufgezeichnet. [41].

Die Dokumentation der gesammelten Aktivitäten in COPRA werden exportiert, sodass Therapieverfahren und Maßnahmen durch das Controlling ermittelt werden [41].

COPRA nutzt etablierte IT-Technologien, wie Microsoft SQL Server und .NET Framework, ist skalierbar, virtualisierbar und bietet Freiheiten bei der Gestaltung der Infrastruktur, sodass es für die IT-Abteilungen attraktiv ist [41].

Durch die Freiheiten bei der Gestaltung der Infrastruktur von COPRA ist ein multidimensionales Datenmodell wie in der Abbildung 2.6 möglich.

Das Schema der Abbildung 2.6 stellt nur das Datenmodell des Bereichs für die Durchführung dieses Projekts dar, denn das COPRA-System beinhaltet ein Schema mit mehr als 260 Tabellen insgesamt im Release 44.0 der Version 1 von 20.10.2015 [43]. Dieses Modell sollte aber nicht mit einem Sternschema in einem DWH (2.6.2) verwechselt werden, denn bei COPRA stellt die Tabelle `co6_medic_data_patient` mit der Basisinformation der behandelnden Personen, in der DWH-Theorie (2.6), eine Dimension dar, und die weiteren Tabellen sammeln Metadaten und die Ergebnisse der Messungen der medizinischen Geräte (Werttabellen), also die Fakten, und beinhalten auch die Hauptschlüssel von `co6_medic_data_patient` als Fremdschlüssel.

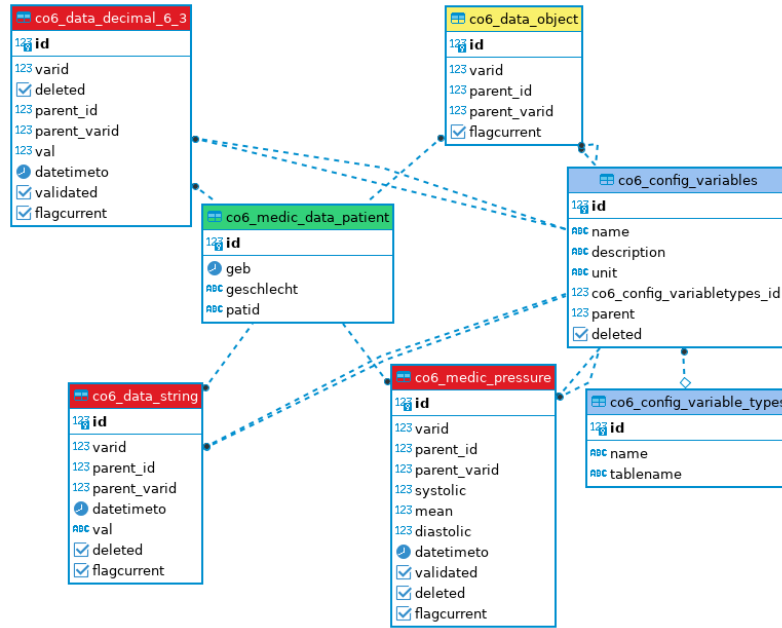


Abbildung 2.6: Datenmodell des für diese Arbeit benutzten Teils vom COPRA-System an der Universitätsmedizin Mainz. Die Tabelle `co6_medic_data_patient` beinhaltet die Information der behandelnden Personen. Die Tabellen `co6_data_decimal_6_3`, `co6_medic_data_pressure` und `co6_data_string` speichern die Ergebnisse der Messungen oder Techniken. Die Tabelle `co6_data_object` beinhaltet die Schlüssel aller Objekte, z. B. Patienten und abstrakte Elemente, wie die Arztbriefe oder Scores. Die Tabellen `co6_config_variables` und `co6_config_variable_types` beinhalten die Metadaten der gespeicherten Informationen in COPRA.

Das Teil der COPRA-DB ist zugleich eine der Datenquelle im Staging Bereich des DWH im DIZ an der Universitätsmedizin Mainz (Abbildung 4.1).

2.6 Data Warehouse

Ein Data Warehouse (DWH) ist „eine fachlich orientierte, integrierte, nicht volatile und in der Zeit veränderliche Sammlung von Daten zur Unterstützung von Entscheidungen“ [44]. Die Eigenschaft der Fachorientierung eines DWH ermöglicht auch die Nutzung solche Systeme für Forschungszwecke [45]. Ein DWH ist auch „eine physische Datenbank, die eine integrierte Sicht auf beliebige Daten zu Analysezwecken ermöglicht“ [46]. Die folgenden Eigenschaften lassen sich aus der Definition ableiten [47].

- Fachorientierung: Die Datenstruktur im DWH wird für die Unterstützung von klinischen Problemstellungen und Entscheidungen optimiert und somit für die Berechnung von Kennzahlen und deren Zuordnung
- Integration: Daten aus diversen Datenquellen werden in getrennte DBs integriert, z. B. die Informationen der verschiedenen Quellsysteme in einem Krankenhaus werden in unterschiedlichen DBs des Staging Bereichs eines DWH gesammelt
- Beständigkeit: Die geladenen Daten im DWH bleiben unverändert
- Historisierung: Die über lange Zeit gespeicherten Daten ermöglichen die Erkennung von Entwicklungen und Trends in den Daten

Die Anwendungsbereiche von DWHs in klinischen Einrichtungen umfasst die Bereitstellung von Informationen, komplexe und flexible Datenanalyse, die Entwicklung von Forschungsprojekten und die Unterstützung der Planungsprozesse [47].

Die Architektur eines DWH besteht aus fünf Ebenen [48]. Die Ebene der Datenquellen umfasst die internen und externen Datenquellen mit relevanten Daten [46]. In der Datenerfassungsebene werden diese Daten mit Hilfe von Extraction, Transformation, Load (ETL)-Prozessen (2.6.1) für das Laden in die zentrale DB-System (Datenhaltungsebene) extrahiert, aufbereitet und konvertiert [44]. Diese zentrale DB ist in mehreren DBs ausgeteilt, und jede davon stellt eine bestimmte Datenquelle dar [46]. In der Datenbereitstellungsebene werden die gespeicherten Daten für multidimensionale Analysen aufbereitet [48]. Auf der Präsentationsebene befinden sich Anwendungsprogramme mit Zugriff auf die aufbereiteten Daten für die Analyse [49].

2.6.1 ETL

Die Extraction, Transformation, Load (ETL)-Prozesse oder Erfassungsprozesse in einem DWH dienen der Aufbereitung und dem Laden der Daten in die DBs des DWH. Diese Prozesse sind je nach Datenmenge und Design des DWH sehr zeitintensiv [48, 49].

Zuerst werden bei der Datenextraktion die Daten aus den verschiedenen Datenquellen extrahiert und in einen Zwischenspeicher übertragen, dem sogenannten Staging Bereich [49]. In diesem Prozess werden die ersten Umwandlungen, wie die Pseudonymisierung von personenbezogenen Daten und die Konvertierung von Datentypen zwischen dem Quellsystem und dem DWH, vorgenommen [48].

Nach der Datenextraktion werden die Daten transformiert. Ziel der Transformation ist eine hohe Qualität der Daten zu gewährleisten. Damit werden die extrahierten Daten bereinigt, harmonisiert und zusammengeführt. Bei der Transformation werden Fehler und Inkonsistenzen beseitigt. Ein weiterer Aspekt der Datentransformation ist die Erstellung von Metadaten und künstlichen Schlüsseln zur Erleichterung der Referenzen auf die Datensätze im DWH, denn in manchen Fällen werden die Schlüssel der Quellsysteme nicht verwendet, denn zwei oder mehrere Quellsystemen können dasselbe Schlüsselschema haben [48, 49], z. B. die Nutzung von fortlaufenden Nummern als Primärschlüssel für die Identifikation von Diagnosen in verschiedenen Datenquellen.

Die Ladephase ist die Letzte im ETL-Prozess. Damit werden die transformierten Daten physikalisch in die Zieltabellen des DWH überführt [46, 49].

Ein Beispiel eines ETL-Prozesses ist in der Abbildung 4.1 im Kapitel „Realisierung“ (4) dieses Projekts dargestellt.

2.6.2 Datenmodell eines DWH

Das Datenmodell eines DWH orientiert sich an Sachverhalten eines Unternehmens, und somit ist ein DWH themenorientiert und bietet eine aggregierte Sicht auf das Geschehen [46].

Das Datenmodell eines DWH ist multidimensional und somit weit verbreitet für die Darstellung analytischer Daten, außerdem ist dieses Modell relativ einfach und vereinfacht die Modellierung der physikalischen DB eines DWH [48]. Andere Vorteile dieser Modellierungsart sind die Performanz und die verständliche Darstellung der Daten für wissenschaftliche Fragestellungen [49].

Das multidimensionale Modell besteht aus Dimensionen und Fakten [44]. Die Dimensionen (Dimensionstabellen) sind Konzepte oder Hierarchien, um die Daten zu kategorisieren. Die Fakten (Faktentabellen) sind Beobachtungen, Berechnungen oder Messungen [49]. Der Hauptschlüssel der Faktentabellen ist meist ein künstlicher Schlüssel, der die Kombination der Hauptschlüssel der Dimensionstabellen enthält [48, 49].

Die Daten in einem multidimensionalen Datenmodell werden in einer physikalischen DB eines Relational Database Management System (RDMS), wie PostgreSQL oder Microsoft SQL Server, meist in einem Sternschema gespeichert und dargestellt [44]. Seinen Namen verdankt das Schema seiner Ähnlichkeit mit einem Stern, wie in der Abbildung 2.7 dargestellt wird [49].



Abbildung 2.7: Sternschema. Die Dimensionstabellen beinhalten Kategorien. Die Faktentabelle beinhaltet die Hauptschlüssel der Dimensionstabellen und die Ergebnisse von Analysen. „PK_“ stellt die Hauptschlüssel der Tabellen dar.

2.7 Data Mapping

Data Mapping oder Datenzuordnung ist die Herstellung der Verbindung zwischen den Attributen zweier unterschiedlicher Datenmodelle (Quelle und Ziel) [50]. Dieses Verfahren wird angewendet, um unterschiedliche Datenquellen zu integrieren und sinnvoll zu nutzen [51]. Die Datenzuordnung ist entscheidend für weitere Schritte in Projekten bei denen zwei oder mehrere Datenmodelle eine Rolle spielen [50]. Das Ziel der Datenzuordnung ist die Gewährleistung des Transports von Information von der Quelle zum Ziel mit so wenig Datenverlust wie möglich [51]. Für den Erfolg des Data Mappings sollte man genau wissen, wie die Daten vom Quellsystem zum Zielsystem fließen [50].

In der Gesundheitsbranche hilft die Datenzuordnung durch die Abgleichung von Daten zwischen Quell- und Zielsystemen die Interoperabilität für die elektronische Patienten Akte zu erreichen [5, 51]. Allerdings hilft die Datenzuordnung auch dem medizinischen Personal, Informationen der behan-

delnden Personen auszutauschen und Gesundheitsdaten aus den verschiedenen Datenquellen zu kombinieren [51].

Dadurch dass der Staging Bereich eines DWH der Zwischenspeicher ist, in dem die Daten aus den Quellsystemen für die spätere Transformationen gelagert werden, ist dieser Bereich sehr entscheidend für die Datenzuordnung, weil das Data Mapping die inhaltliche Information und Datenstruktur verschiedener Tabellen aus dem Quellsystem benötigt [50, 52].

Für die Durchführung eines Data Mappings sind drei Elemente entscheidend: das Logical Data Model (LDM), das Physical Data Model (PDM) und die fachlichen Spezialisten [50]. Einerseits liefert das LDM die Details der Strukturdefinition der Datenquelle [50, 53]. Andererseits beschreibt das PDM die Spezifikationen der Implementierung des LDM [53]. Die fachlichen Spezialisten besitzen das Fachwissen über eine bestimmte Thematik und stehen somit bei Fragen und Problemen zur Verfügung [54].

Es gibt drei Data Mapping-Haupttechniken [51]:

1. Manuelle Datenzuordnung: Manuelle Kodierung der Datenquellen oder manuelle Zuordnung des Zielschemas
2. Schemazuordnung: Halbautomatischer Prozess bei dem eine Beziehung zwischen Quell- und Zielschema hergestellt wird und die Verbindungen geprüft und gegebenenfalls angepasst werden
3. Vollautomatische Datenzuordnung: Diese Technik bietet eine bequeme, einfache und effiziente, meist codefreie, Benutzeroberfläche

Das Data Mapping wird in verschiedenen Szenarien angewendet. Eines davon ist, wie bei diesem Projekt, der elektronische Datenaustausch von Gesundheitsdaten, bei dem die Quelldaten in verschiedene Formate konvertiert werden sollen [51].

Das Data Mapping besteht aus sechs Schritten [55].

1. Definition der Daten für die Verschiebung - Tabellen, Felder und Datenformat der Felder werden definiert
2. Zuweisung der Quell- und Zielfelder
3. Transformation - Programmierung der Transformationsregeln
4. Test - Test mit Beispieldaten aus der Quelle, um mögliche Anpassungen vorzunehmen
5. Implementierung

6. Migration oder Integration der Daten

In dieser Masterarbeit werden zum Teil einige Transformationsregeln definiert, und programmiert.

Dadurch, dass Data Mapping eine dynamische Schnittstelle ist, sollte es regelmäßig gewartet und aktualisiert werden [55].

Dieses Projekt befasst sich mit dem manuellen Data Mapping für den Datenaustausch von Biosignaldaten aus der COPRA-Instanz des Staging Bereichs eines DWH, als Quellsystem, zu einem Zielsystem in FHIR. Das manuelle Data Mapping wurde ausgewählt, denn bis jetzt existieren noch keine vorherigen Projekte - weder an der Universitätsmedizin Mainz noch an anderen MII Standorte, welche sich mit dem Data Mapping von Biosignaldaten aus COPRA mit FHIR-Profilen befassen.

2.8 Pattern Matching

Bei der String Suche oder dem Pattern Matching ist das Problem, alle Elemente einer Zeichenkette x der Länge p (Pattern) in einer anderen Zeichenkette t (Text) der Länge n zu lokalisieren [56]. Eine weitere hilfreiche Definition, um das Pattern Matching in der Praktik zu verstehen, ist die Bezeichnung „Teilwort“. „Ein Wort T heißt Teilwort des Wortes W , wenn es Worte U , V gibt, so dass gilt: $W = UTV$ “ [57]. Genau gesagt, ist das Pattern Matching der Prozess ein Teilwort eines Wortes zu erkennen.

Das Pattern Matching wird in verschiedenen Szenarien, wie die Textverarbeitung [58], Netzwerksysteme oder DB-Schemata, verwendet [59]. Einige der Anwendungsbereiche sind Texteditoren in Computern für die Syntaxprüfung von Sprachen oder Rechtschreibprüfung, Programmierung von Datenbankabfragen, in Bioinformatik für den Abgleich von DNA- oder Eiweißsequenzen, Entwicklung von Systemen zur Erkennung von Eindringlingen in Netzwerke, digitale Bibliotheken, Suchmaschinen und viele weitere Anwendungen [60, 59].

Einige der Methoden um Zeichenketten oder Mustern zu finden sind die unscharfe Suche (fuzzy search) (2.8.1) und die Anwendung von Regular Expressions (Regexs) (2.8.2).

Manche Programmiersprachen, wie Python und RDMS, wie PostgreSQL, besitzen Features für die unscharfe Suche und Regexs, um einen Text nach Mustern zu durchsuchen [62, 61].

2.8.1 Unscharfe Suche

Bei der unscharfen Suche, fuzzy-Suche oder fuzzy search wird eine Ähnlichkeit-Funktion benutzt. Diese Funktion bildet ein Paar Zeichenfolgen s und t auf eine reelle Zahl r ab, wobei ein größerer Wert von r eine größere Ähnlichkeit zwischen s und t anzeigt [63]. Eine der bekannten Funktionen ist die „Levenshtein distance“. Diese Funktion stellt die Kosten für die Transformation eines Strings in einem anderen dar. Die erlaubten Operationen für die Transformationen sind die Einführung neuer Charakteren, das Löschen und die Substitution von Zeichen [63]. Eine Variante der „Levenshtein distance“ betrachtet auch die Transposition von Charakteren [64].

2.8.2 Regex

Die Regular Expressions (Regexs) oder reguläre Ausdrücke sind eine Art Sprache, die in der Informatik für diverse Zwecke angewendet werden kann, insbesondere bei der Bearbeitung von Texten oder bei der Suche von bestimmten Mustern in einem Text [60]. Ein regulärer Ausdruck wird als eine Expression definiert, die eine Menge von Zeichenketten „reguläre Sprache“ oder eine Menge von geordneten Paaren von Zeichenketten „reguläre Beziehung“ beschreibt [65]. Sie sind in einer Vielzahl von Programmiersprachen und IT-Anwendungen verfügbar. Ein Regex wird mit einem Muster von Symbolen erstellt, das sogenannte Metazeichen, die zur Definition des Musters dienen und auch syntaktische Regeln besitzen [66].

- `\d`: Platzhalter für eine Ziffer, z. B. `5`
- `\w`: Platzhalter für ein alphanumerisches Zeichen oder Unterstrich „_“, z. B. `G`
- `\W`: Platzhalter für ein Zeichen, dass keine alphanumerisches Zeichen oder Unterstrich ist, z. B. `?`
- `[...]`: Definiert ein Set von Zeichen, z. B. `[acd]`
- `|`: Trennung von zwei oder mehreren Alternativen, z. B. `x | y`
- `+`: Mindestens einmal
- `{n}`: Genau `n`-mal
- `{min, max}`: Mindestens `min`-mal und maximal `max`-mal
- `{min, }`: Mindestens `min`-mal

- {, max}: Maximal **max**-mal
- (...): Gruppierung

Das folgende Beispiel zeigt die Definition eines Regex für die Erkennung von Zeichenketten, wie die Maßeinheiten cm[Hg] oder mm[H2O]. Der Regex „[cm]m\WH(g|2O)\W“ definiert eine Zeichenkette mit „c“ oder „m“ gefolgt von „m“. Nach diesem Charakter muss ein nicht alphanumerisches Zeichen platziert werden, dieses ist von einem „H“ und einer Gruppe von Charakteren mit entweder „g“ oder „2O“ gefolgt. Nach der Gruppierung muss sich ein nicht-alphanumerisches Zeichen befinden.

Methode

Um das Ziel dieses Projekts zu erreichen, werden die im PDMS der Universitätsmedizin Mainz gespeicherten Biosignaldaten den FHIR-Profilen des Erweiterungsmoduls „Intensivmedizin“ des Kerndatensatzes der MII zugeordnet und für die Überführung in FHIR-Ressourcen bereitgestellt. Dazu werden etablierte IT-Werkzeuge, wie Structued Query Language (SQL), Data Mapping, Pattern Matching und ETL-Prozesse angewendet.

Für die Durchführung dieses Projekts an der Universitätsmedizin Mainz werden die Parameter der FHIR-Profile des Erweiterungsmoduls „ICU“ zusammen mit den, in der COPRA-Instanz des Staging Bereichs des DWH des DIZ, gespeicherten Konfigurationsvariablen in eine DB importiert. Diese Variablen stellen unter anderem die Messungen und Beobachtungen der Bioparameter dar. In dieser DB werden die notwendigen Schritte für die Zuordnung der Konfigurationsvariablen mit den FHIR-Profilen durchgeführt.

Nach dieser Zuordnung wird der resultierende validierte Datensatz in der COPRA-Instanz des Staging Bereichs des DWH für die Bereitstellung der Daten für die Implementierung und Automatisierung des Prozesses des Exports der Biosignaldaten in FHIR-Ressourcen gespeichert.

3.1 Schritte für die Durchführung

Für die Umsetzung dieser Arbeit wird eine DB in PostgreSQL entwickelt, um die Information der FHIR-Profile und Konfigurationsparameter der COPRA-Instanz des Staging Bereichs des DWH des DIZ an der Universitätsmedizin Mainz zu speichern, und einige der Schritte des Data Mappings durchzuführen.

Die FHIR-Profile des Erweiterungsmoduls „ICU“ werden analysiert, und die allgemeinen Parameter der Profile werden in der entwickelten DB gespeichert. Einige dieser Parameter werden als Hilfsmittel für die Datenzuordnung angewendet. Andere hingegen spiegeln die Attribute der Biosignaldaten in COPRA wider.

Nach der Speicherung der Profile wird die Struktur der Tabellen von COPRA im Staging Bereich des DWH analysiert. Diese beinhalten die Metadaten, wie Konfigurationsvariablen und Datentypen, und die Werte der Biosignaldaten mit den Attributen für das Data Mapping mit den FHIR-Profilen.

Die Voraussetzungen für die Auswahl der Konfigurationsvariablen für die weitere Durchführung des Projekts werden definiert. Die ausgewählten Konfigurationsvariablen sollen patienten- oder fallbezogene sein, und mindestens 1000 validierten, aktuellen und nicht gelöschten Datensätzen in den Werttabellen beinhalten (3.2.1).

Nach der Auswahl werden die entscheidenden Attribute der Konfigurationsvariablen in die vorher entwickelten PostgreSQL DB importiert. Diese Konfigurationsvariablen beinhalten die Namen von Beobachtungen und Messungen. Diese solle den „Procedure“- und „Observation“-Profilen entsprechen.

Mit den gespeicherten Informationen der FHIR-Profile und den Konfigurationsvariablen in der DB und der analysierten Datenstruktur von COPRA wird das manuelle Data Mapping durchgeführt. Von diesem Prozess werden in diesem Projekt drei Schritte umgesetzt: die Datendefinition, Zuweisung der Quell- und Zielfelder, und die Programmierung der Transformationsregeln.

Der Prozess für die Verlinkung der Konfigurationsvariablen mit den FHIR-Profilen generiert einen Datensatz mit den notwendigen Attributen der Konfigurationsvariablen und der FHIR-Profile. Dieser Datensatz wird validiert.

Für die Harmonisierung der Maßeinheiten beider Systeme werden diese Einheiten verglichen, um Fehler und andere Irregularitäten zu detektieren, und wenn möglich, mit der Kooperation der fachlichen Spezialisten der PDMS-Abteilung und mit Hilfe der Definition von Transformationsregeln solche Probleme zu lösen. Diese Regeln werden im Datensatz der zugeordneten Konfigurationsvariablen mit den FHIR-Profilen eingefügt, sodass zum Schluss die Maßeinheiten beider Systeme harmonisiert sind.

Der Datensatz der zugeordneten Konfigurationsvariablen mit den FHIR-Profilen wird danach in die COPRA-Instanz des Staging Bereichs des DWH angelegt, um weitere Transformationsregeln für die Erzeugung der FHIR-Ressourcen aus den Biosignaldaten von COPRA zu definieren.

Mit dem Datensatz der zugeordneten Konfigurationsvariablen im Staging Bereich des DWH und mit der Identifikation der Felder von Quelle- und Zielsystemen werden SQL-Views im COPRA-Instanz des Staging Bereichs programmiert, um die notwendigen Parameter der verschiedenen Tabellen zusammenzuführen. Diese Views können bei der Implementierung für die Erzeugung der FHIR-Ressourcen angewendet werden.

Ein Flussdiagramm (C) der Ablaufschritte dieses Projekts ist in der Ab-

bildung 3.1 dargestellt.

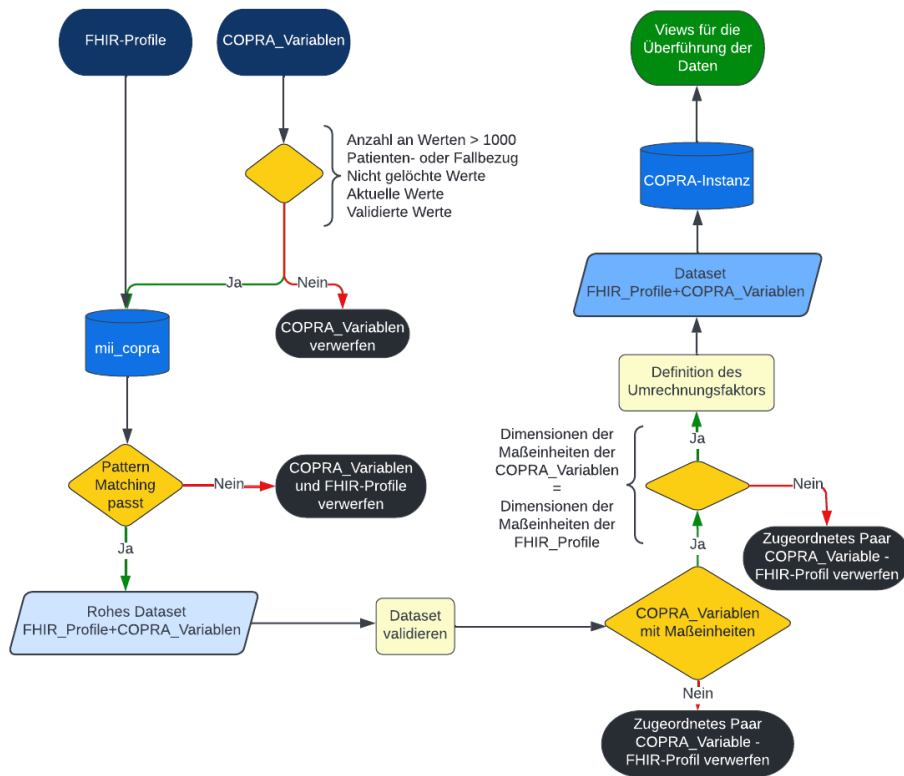


Abbildung 3.1: Flussdiagramm des Ablaufes des Projekts.

3.2 Data Mapping der Biosignaldaten aus COPRA mit den FHIR-Profilen

Nach der Analyse der FHIR-Profilen des Erweiterungsmoduls „Intensivmedizin“ und des COPRA-Datenmodells wurden die Information beider Systeme für die Datendefinition analysiert und ein Pattern Matching-Prozess für die Zuordnung der Konfigurationsvariablen aus COPRA mit den FHIR-Profilen durchgeführt. Während dieses Verfahrens entstand ein Datensatz. Dieser wurde validiert und für die Zuweisung der Quell- und Zielfelder, und die Programmierung der Transformationsregeln benutzt.

3.2.1 Datendefinition

Die Tabellen mit den Attributen der FHIR-Profile (4.2) und die Tabellen von COPRA (4.5) wurden untersucht, um zu erkennen, welche COPRA-Tabellen und Spalten mit den Elementen der FHIR-Profile zu vernetzen sind.

Die Tabelle der FHIR-Profile (Tabelle 4.1) beinhaltet die Information der Elemente der Profile des Moduls „ICU“ des Kerndatensatzes der MII und somit die Datenstruktur und Parameter des Zielsystems.

Das Quellsystem dieser Arbeit sind die Tabellen der COPRA-Instanz des Staging Bereichs des DWH des DIZ an der Universitätsmedizin Mainz. In der Tabelle der Konfigurationsvariablen (Tabelle 4.6) sind die am Standort angegebenen COPRA-Namen und Beschreibungen der Messungen oder angewandten Techniken zusammen mit deren Maßeinheiten zu finden. Die Tabelle der Typen der Konfigurationsvariablen (Tabelle 4.7) beinhaltet die Datentypen der Konfigurationsvariablen und die Namen der Werttabellen, in den die Werte der Biosignaldaten gespeichert sind. Eine andere wichtige Tabelle in COPRA ist die mit den Basisdaten der behandelnden Personen (Tabelle 4.8). Die Werte und Zeitangaben der Messungen befinden sich in den Werttabellen (Tabelle 4.10, Tabelle 4.11).

Für das Weiterlaufen des Projekts wurden nur die Konfigurationsvariablen ausgewählt, die einen Patientenbezug oder Fallbezug besitzen, mit mindestens 1000 validierten, aktuellen und nicht gelöschten Datensätzen im COPRA-System, denn unter dieser Gruppe befinden sich die Konfigurationsvariablen, die den Biosignaldaten entsprechen.

3.2.2 Zuweisung der Quell- und Zielfelder

Nach der Definition der Daten wurden die Quell- und Zielfelder zugewiesen. Um diese Zuweisung zu erzielen wurde ein Pattern Matching durchgeführt, um ähnliche Muster zwischen den Parametern der FHIR-Profile und bestimmten Attributen der Konfigurationsvariablen zu erkennen.

Zu Beginn wurden die Namen der FHIR-Profile zusammen mit den Namen und Beschreibungen der Konfigurationsvariablen mit einer unscharfen Suche mit Hilfe von der Bibliothek „thefuzz“ in Python analysiert. Diese unscharfe Suche liefert einen Datensatz mit Paaren von Profilen und Konfigurationsvariablen. Anschließend wurden SQL-Abfragen mit Regexp für jedes Profil entwickelt, um falsch zugeordnete Konfigurationsvariablen herauszunehmen und nicht erkannte Konfigurationsvariablen wahrzunehmen. Die Entscheidung der Nutzung von Regexp statt des SQL-Befehls „LIKE“ basiert auf der Tatsache, dass die Regexp eine breite Palette an Möglichkeiten bietet, um komplexe Muster zu definieren (2.8.2).

Dieser Prozess generierte einen Datensatz mit den zugeordneten Paaren: Konfigurationsvariable - FHIR-Profil. Dieser Datensatz beinhaltet die wichtigsten Attribute beider Systeme und wurde mit Hilfe der Spezialisten der PDMS-Abteilung validiert.

Mit dem validierten Datensatz wurden die Attribute der FHIR-Profile mit den Felder der Tabellen von COPRA zugewiesen, und die Transformationsregel wurden definiert und programmiert.

3.2.3 Transformationsregeln

Für die Definition und Programmierung der Transformationsregeln wurden an erster Stelle die Maßeinheiten beider Systeme im erzeugten Datensatz nach der Validierung verglichen, um Unregelmäßigkeiten bei diesem Attribut zu detektieren und bestimmte Umwandlungen für die Harmonisierung der Maßeinheiten vorzunehmen. Die Umrechnungen wurden im generierten Datensatz des Pattern Matchings eingefügt.

Der resultierende Datensatz wurde in die COPRA-Instanz des Staging Bereichs des DWH importiert. Mit diesem Datensatz in der COPRA-Instanz, zusammen mit den definierten Spalten mit den Parametern der Biosignaldaten in den Werttabellen, den Spalten mit den Attributen der behandelnden Personen, und Metadaten, wurden weitere Transformationsregeln für die Überführung der Biosignaldaten aus COPRA in FHIR programmiert.

3.3 Analyse für die Harmonisierung der Maßeinheiten

Bei der Programmierung der Transformationsregeln ist ein wichtiger Aspekt zu berücksichtigen: die Maßeinheiten beider Systeme, weil sie harmonisiert werden müssen.

Um die Harmonisierung der Maßeinheiten beider Systeme zu erreichen, wurden die Schreibweise und die Dimensionen der physikalischen Größen der Maßeinheiten analysiert. Konfigurationsvariablen bei denen die Dimensionen der Einheiten nicht mit den von den FHIR-Profilen übereinstimmen oder die Maßeinheiten im COPRA-System nicht dokumentiert sind, wurden nicht berücksichtigt. Werte in den Werttabellen mit Maßeinheiten derselben Dimensionen, wie bei den FHIR-Profilen, aber mit anderen Untereinheiten, wurden umgerechnet.

Die Maßeinheiten der FHIR-Profile wurden bei dem Import in der DB gleichzeitig analysiert, und die Anmerkungen, wie unterschiedliche Schreib-

weise derselben Maßeinheit oder Untereinheiten zwischen Profilen, wurden als Issue auf der Webseite von SIMPLIFIER des Moduls „ICU“ gemeldet.

Die Konfigurationsvariablen ohne Maßeinheiten wurden den Spezialisten der PDMS-Abteilung gesendet, denn manche Konfigurationsvariablen beinhalten die Maßeinheiten am Frontend von COPRA und nicht in der DB des Systems.

Für den Vergleich zwischen den Maßeinheiten der Konfigurationsvariablen von COPRA und der FHIR-Profilen in der DB, wurde eine SQL-Abfrage mit integrierten Regexp programmiert. Dieser Schritt ist nicht nur für den Vergleich der Maßeinheiten zwischen beiden Systemen notwendig, sondern auch für die Erkennung von Unregelmäßigkeiten im COPRA-System.

Am Ende dieser Analyse wurde der Datensatz des Pattern Matchings gefiltert, und die zugeordneten Paare von Konfigurationsvariablen und FHIR-Profilen mit Problemen bezüglich der Maßeinheiten wurden herausgenommen. Dieser neue Datensatz wurde in einer neuen Tabelle (4.9) mit einer neuen Spalte für die Umrechnung der Maßeinheiten COPRA kopiert, sodass die Maßeinheiten der Systeme zum Schluss harmonisiert sind. Dieser Datensatz wird wiederum in die COPRA-Instanz des Staging Bereichs des DWH importiert, um weitere Schritte des Data Mappings durchzuführen.

3.4 Vorbereitung der Überführung der Daten aus COPRA in FHIR

Der erzeugte Datensatz nach der Harmonisierung der Maßeinheiten wurde in die COPRA-Instanz des Staging Bereichs des DWH importiert, denn der Rest der benötigten Daten für die Erzeugung der FHIR-Ressourcen des Erweiterungsmoduls „Intensivmedizin“ befinden sich in verschiedenen Tabellen dieser Instanz.

Dadurch, dass die notwendigen Daten in COPRA in mehreren Tabellen liegen (Abbildung 2.6), sollten sie hierzu zusammengeführt werden. Für die Zusammenführung der Informationen wurden die genauen Daten von COPRA und den FHIR-Profilen des Datensatzes der Zuordnung spezifiziert. Mit dieser Spezifikation wurden anschließend SQL-Views mit den zusammengeführten Parametern und den Transformationsregeln in der COPRA-Instanz programmiert. Diese Views können in weiteren Prozessen bei der Überführung der Biosignaldaten aus COPRA in FHIR benutzt werden.

Realisierung

Dieses Kapitel befasst sich mit der Realisierung dieses Projekts und die Ereignisse, die währenddessen entstanden sind.

Vor der Realisierung des Projekts ist es wichtig zu wissen, wie die Daten in dem Staging Bereich des DWH des DIZ an der Universitätsmedizin Mainz gespeichert werden, und wie diese Daten danach im DIZ bearbeitet werden (Abbildung 4.1).

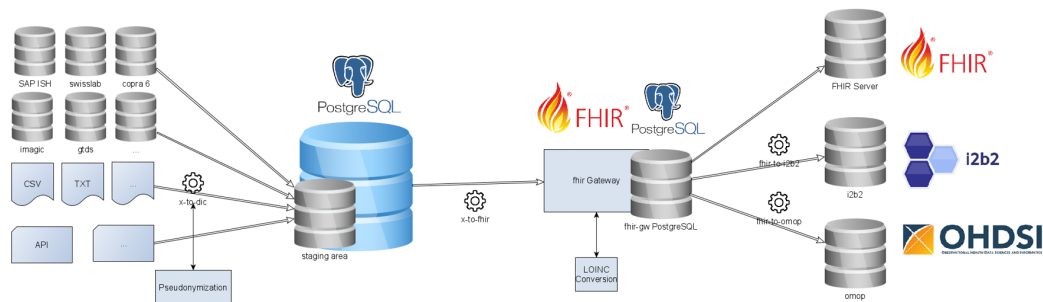


Abbildung 4.1: DIZ-Struktur an der Universitätsmedizin Mainz. Die Daten aus den verschiedenen Quellen in der Klinik werden pseudonymisiert in dem Staging Bereich des DWH zur Bearbeitung gespeichert. Diese bearbeiteten Daten werden in FHIR überführt oder zu anderen DBs transportiert. Am Ende wird die Information mit Hilfe von Werkzeugen, wie i2b2 visualisiert.

Für die Durchführung dieser Masterarbeit wurde mit der Information der COPRA-Instanz des Staging Bereichs des DWH gearbeitet (Abbildung 4.2).

Zu Beginn wurde die DB `mii_copra` entwickelt. In den Tabellen dieser DB wurden die Parameter der FHIR-Profil des Erweiterungsmodul „Intensivmedizin“ zusammen mit den relevanten Konfigurationsvariablen von COPRA importiert. In der DB `mii_copra` wurden auch die Zwischenschritten des Data Mappings realisiert. Als Ergebnis dieses Prozesses entstand der Datensatz mit zugeordneten Konfigurationsvariablen mit FHIR-Profilen. Die Maßeinheiten im Datensatz wurden harmonisiert. Dieser Datensatz wurde danach in

der COPRA-Instanz des Staging Bereichs des DWH importiert. Mit diesem Datensatz als Basis wurden SQL-Views für die Zusammenführung der Daten beider Systeme programmiert.

Ein Diagramm mit den Komponenten und dem Fluss der Daten in diesem Projekt ist in der Abbildung 4.2 dargestellt.

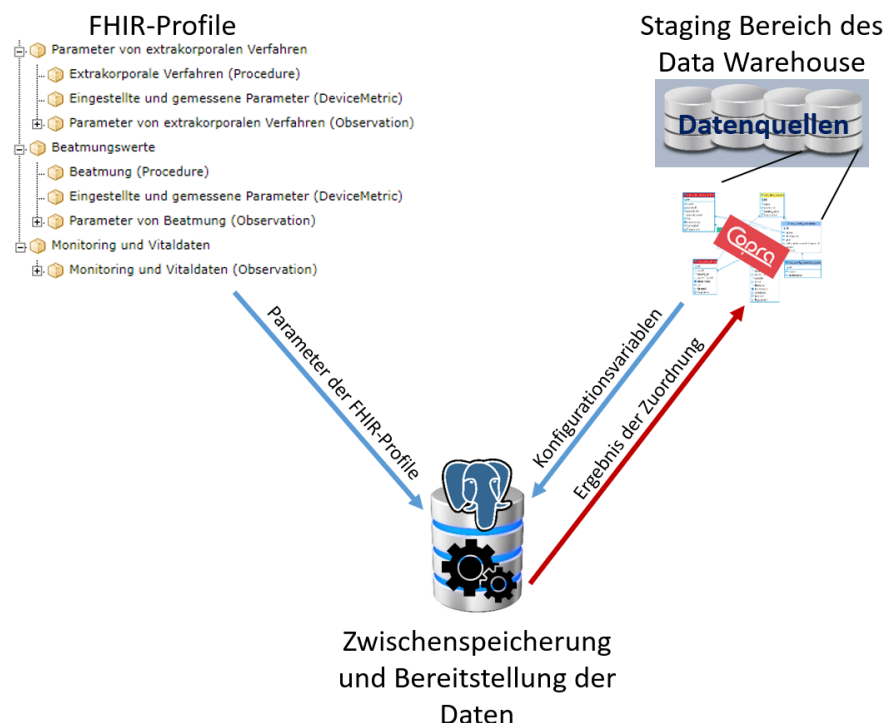


Abbildung 4.2: Komponenten und Fluss der Daten in diesem Projekt.

4.1 PostgreSQL-DB

Für die Entwicklung dieses Projekts wurde die DB `mii_copra` in PostgreSQL erstellt, um Attribute der FHIR-Profile und Parameter der im COPRA gespeicherten Konfigurationsvariablen zu speichern. In dieser DB fanden auch die Bearbeitung und Analyse der Daten statt.

Die DB `mii_copra` wurde in einer Testumgebung mit PostgreSQL 14.2 unter Linux Ubuntu 22.04.1 Long Term Support (LTS) implementiert. Für die Entwicklung der DB und die weiteren SQL-Prozesse wurde die Community Version des freien Open Source Werkzeugs DBeaver benutzt.

In der DB `mii_copra` wurde eine Tabelle angelegt für die Speicherung bestimmter herausgenommener Parameter der FHIR-Profile der MII - Modul

„ICU“. Zwei weitere Tabellen sollen die relevanten Attribute der Metadaten von COPRA speichern. Diese Tabellen mit Metadaten sammeln die Informationen der Konfigurationsvariablen und deren Typen. In der DB wurden auch die SQL-Abfragen für die Zusammenführung der FHIR-Profile mit den Biosignaldaten geschrieben. Schließlich wurden weitere Tabellen angelegt, um die resultierenden Zwischenergebnisse und Datensätze zu speichern, analysieren und visualisieren (Tabelle 4.1, Tabelle 4.14).

4.2 Parameter der FHIR-Profile des Moduls „ICU“

Die Information der FHIR-Profile wurde aus der Webseite Medizininformatik Initiative - Modul ICU - ImplementationGuide herausgenommen.

Obwohl dieses Modul zum Abstimmverfahren von 25.02.2022 bis 08.04.2022 freigegeben wurde, befindet es sich noch während der Entwicklung dieses Projekts in der ersten stabilen Ballot-Version. Das bedeutet für diese Arbeit, dass Unregelmäßigkeiten in den Profilen detektiert werden können, und als Issue gemeldet werden.

Die Parameter je Profil an der Webseite: Name, Typ, Attribute der Maßeinheit, Daten der Codesysteme und vorhandene Metrik des Geräts, wurden in die Tabelle `mii_icu` der DB `mii_copra` eingefügt.

Die meisten FHIR-Profile unter der Kategorie „Observation“ besitzen eine ähnliche Struktur, sodass der Import in die Tabelle `mii_icu` erleichtert wird. Die Ausnahmen bilden die Profile der Gruppe „Blutdruck Generisch“ zur Speicherung der Blutdruckmessungen, z. B. „Linksatrialer Druck“ (2.2.1). In solchen Fällen wurden auch die semantischen Annotationen der systolischen, mittleren und diastolischen Attribute in der Tabelle `mii_icu` registriert.

Die Struktur der Tabelle `mii_icu` zur Speicherung der Elemente der FHIR-Profile ist in der Tabelle 4.1 dargestellt.

Tabelle 4.1: Struktur der Tabelle `mii_icu`. Im Feld Spalte befinden sich die gegebenen Namen der FHIR-Elemente der Profile. Der Datentyp ist der Typ mit dem die Daten in der Tabelle `mii_icu` gespeichert werden. Die Spalte Information speichert die Beschreibung des Elements.

Spalte	Datentyp	Information
<code>profile_id</code>	int	Generierter numerischer Identifikator des zugeordneten FHIR-Profils
<code>profile_name</code>	varchar	Name des FHIR-Profils, z. B. Atemzugvolumen-Waehrend-Beatmung
<code>category_coding_system</code>	text	URL der Kategorie „Observation“. Dies kann ein SNOMED CT-URL sein
<code>category_coding_code</code>	varchar	„vital-signs“ oder SNOMED CT-ID
<code>code_coding_system_snomed</code>	text	http://snomed.info/sct
<code>code_coding_code_snomed</code>	varchar	SNOMED CT des FHIR-Profils, z. B. 250874002
<code>code_coding_system_loinc</code>	text	http://loinc.org
<code>code_coding_code_loinc</code>	varchar	LOINC-Code des FHIR-Profils, z. B. 76222-9
<code>code_coding_system_ieee</code>	text	urn:iso:std:iso:11073:10101
<code>code_coding_code_ieee</code>	varchar	ISO/IEEE 11073-10101™-Schlüssel des FHIR-Profils, z. B. 151980
<code>valuequantity_system</code>	text	http://unitsofmeasure.org
<code>valuequantity_code</code>	varchar	Maßeinheiten im FHIR-Profil, z. B. mL
<code>device_reference</code>	text	Zuweisung zur Art der Prozedur (gemessen, eingestellt oder erhoben). Diese Prozedur befindet sich in den Profilen der Typ „DeviceMetric“

code_systolic _coding_system _snomed	text	http://snomed.info/sct
code_systolic _coding_code _snomed	varchar	SNOMED CT-ID der systolischen Blutdruckmessung, z. B. 271649006
code_systolic _coding_system _loinc	text	http://loinc.org
code_systolic _coding_code_loinc	varchar	LOINC-Code der systolischen Blutdruckmessung, z. B. 8406-1
code_systolic _coding_system _ieee	text	urn:iso:std:iso:11073:10101
code_systolic _coding_code_ieee	varchar	ISO/IEEE 11073-10101™-Schlüssel der systolischen Blutdruckmessung
code_mean _coding_system _snomed	text	http://snomed.info/sct
code_mean _coding_code _snomed	varchar	SNOMED CT-ID der mittleren Blutdruckmessung, z. B. 6797001
code_mean _coding_system _loinc	text	http://loinc.org
code_mean _coding_code_loinc	varchar	LOINC-Code der mittleren Blutdruckmessung, z. B. 8478-0
code_mean _coding_system _ieee	text	urn:iso:std:iso:11073:10101
code_mean _coding_code_ieee	varchar	ISO/IEEE 11073-10101™-Schlüssel der mittleren Blutdruckmessung, z. B. 150019
code_diastolic _coding_system _snomed	text	http://snomed.info/sct
code_diastolic _coding_code _snomed	varchar	SNOMED CT-ID der diastolischen Blutdruckmessung, z. B. 271650006

code_diastolic _coding_system _loinc	text	http://loinc.org
code_diastolic _coding_code_loinc	varchar	LOINC-Code der diastolischen Blutdruckmessung, z. B. 8462-4
code_diastolic _coding_system _ieee	text	urn:iso:std:iso:11073:10101
code_diastolic _coding_code_ieee	varchar	ISO/IEEE 11073-10101™-Schlüssel der diastolischen Blutdruckmessung, z. B. 150018
meta_profile	text	URL um das Profil zu identifizieren

4.3 Analyse der FHIR-Profile des MII - Moduls ICU

Bevor die Parameter der FHIR-Profile in die DB importiert wurden, erfolgte eine Analyse der bis dato 80 FHIR-Profile.

In dieser Analyse wurde beobachtet, dass zwei Profile dieselben Namen besitzen, aber unterschiedliche Information beinhalten. Diese sind unter dem Profil-Namen „Linksventrikulärer Schlagvolumenindex“ zu finden. In einem der Profile wird tatsächlich die Information des linksventrikulären Schlagvolumenindex spezifiziert. Das andere Profil hingegen sollte das linksventrikuläre Schlagvolumen definieren. Dieses Ereignis wurde durch das Attribut „code“ beider Profile erkannt. Dieses Attribut spezifiziert die Codesysteme des Profils. An dieser Stelle wurde beobachtet, dass das Profil für den linksventrikulären Schlagvolumenindex den LOINC-Code 76297-1 „Left ventricular Stroke volume index“ besitzt. Das Profil für das linksventrikuläre Schlagvolumen beinhaltet wiederum den LOINC-Code 20562-5 „Left ventricular Stroke volume“. Noch dazu sind die spezifizierten SNOMED CT-IDs und IEEE-Schlüssel beider Profile im Attribut „code“ auch unterschiedlich. Diese Problematik war in der SIMPLIFIER-Webseite des Moduls vorher korrigiert. Aus diesem Grund wurde diese Anmerkung als Issue nicht gemeldet.

Nach der durchgeführten Analyse der 80 FHIR-Profile, wurden die FHIR-Elemente der Profile in der Tabelle `mii_icu` der DB `mii_copra` geschrieben. Mit diesen Daten in der DB wurde die Einteilung der Profile analysiert.

Die Tabelle 4.2 zeigt die Anzahl an Profilen je Kategorie im Modul „ICU“. Die meist repräsentierten FHIR-Profile, nämlich 76, gehören zu der Kategorie „Observation“.

Tabelle 4.2: FHIR-Profile im Modul „ICU“.

Anzahl der Profile	Profiltyp
76	Observation
2	Procedure
2	DeviceMetric
80	Gesamt

In den logischen Tabellen der FHIR-Profile sind die standardisierten Codesysteme der Biosignaldaten im Attribut „code“ registriert. Dieses Attribut ist charakteristisch für Profile der Typen „Observation“ und „Procedure“. Die Distribution der Codesysteme in der FHIR-Profile ist in der Tabelle 4.3 dargestellt. Das meist repräsentierte Codesystem ist LOINC mit 62 Profilen. Aus diesem Grund wurde dieses Codesystem als zusätzlicher Parameter für das Pattern Matching angewendet. Von den 78 FHIR-Profilen unter der Kategorien „Observation“ und „Procedure“ besitzen fünf kein Codesystem.

Tabelle 4.3: Distribution der Codesysteme der FHIR-Profile im Modul „ICU“.

Anzahl der Profile	Codesystem
62	LOINC
54	SNOMED CT
36	ISO/IEEE 11073-10101™
5	-

Die Profile der Typen „Observation“, die keine Codesysteme beinhalten (Tabelle 4.4), gehören zu den generischen Profilen. Diese sind nicht für die direkte Umsetzung konzipiert, sondern zur Modellierung der Daten. Die „Procedure“-Profile beinhalten keine Codesysteme, weil diese Information im Basismodul „Prozedur“ der Kerndatensatz der MII sich befindet. Die Profile unter der Kategorie „DeviceMetric“ besitzen auch keine Codesysteme, denn diese Profile beschreiben die Messung oder Einstellungen des Geräts (2.2.1).

Tabelle 4.4: „Observation“-Profile ohne Codesystem.

FHIR-Profil	Profiltyp
Parameter von extrakorporalen Verfahren	Observation
Monitoring und Vitaldaten	Observation
Parameter von Beatmung	Observation

Ein weiterer wichtiger Aspekt ist die Analyse der Maßeinheiten der 76 Profile des Typs „Observation“. An dieser Stelle wurden 73 Profile mit definierten und drei ohne definierte Maßeinheiten im Modul gefunden. Diese letzten drei Profile sind dieselben „Observation“-Profile ohne Codesysteme (Tabelle 4.4).

Eine andere Beobachtung hinsichtlich der Maßeinheiten mancher Profile ist, dass die Schreibweise einer Maßeinheit oder deren Untereinheiten bei verschiedenen Profilen nicht immer dieselbe war. Diese Unregelmäßigkeit wurde an die SIMPLIFIER-Webseite des Moduls gemeldet (Issue #2083), und von den Autoren und Entwicklern des Moduls korrigiert. Nach der Korrektur wurden die Schreibweisen der Maßeinheiten der FHIR-Profile in der Spalte `valuequantity_code` der Tabelle `mii_icu` geändert.

Ein Detail bei der Benennung der Profile ist, dass die Namen von 26 Profilen Umlaute beinhalten. Bei sechs Profilnamen hingegen wurden die Umlaute zusammen mit dem Buchstabe Eszett „ß“ vermieden. Die Tabelle 4.5 zeigt einige Beispiele davon. Dieses Phänomen wurde auch als Issue gemeldet (Issue #2394), denn dieselben FHIR-Profile in der ImplementationGuide der SIMPLIFIER-Webseite beinhalten ebenso Umlaute oder das Eszett im Namen.

Tabelle 4.5: An- und Abwesenheit von Umlauten in den Profilnamen

Namen mit Umlaut
Hämodialyse Blutfluss
Zeitverhältnis-Ein-Ausatmung
Namen ohne Umlaut und ohne Eszett
Spontanes-Mechanisches-Atemzugvolumen-Während-Beatmung
Körpergröße

4.4 Angepasste Tabellen von COPRA

Das Datenmodell der COPRA-Instanz des Staging Bereichs (Abbildung 2.6) zusammen mit dem Inhalt der Tabellen dieser COPRA-Instanz wurden analysiert, denn in COPRA befinden sich die Daten und Metadaten der Bioparameter.

Die Tabellen der COPRA-Instanz beinhalten nicht nur die notwendigen Parameter für die Entwicklung dieses Projekts, sondern auch administrative Attribute, die nicht für die Zwecke dieser Arbeit relevant sind. Aus diesem Grund werden nur die relevanten Spalten für das Data Mapping betrach-

tet. Infolgedessen wurden gleichnamige angepasste Tabellen für die Metadaten (`co6_config_variables` und `co6_config_variable_types`) in der DB `mii_copra` zur Bearbeitung der Daten codiert.

Die Tabelle `co6_config_variables` beschreibt die Entitäten im COPRA-System und somit beinhaltet diese Tabelle die Namen und Beschreibungen der Biosignalparameter. In dieser Tabelle wurden nur die Konfigurationsvariablen mit den notwendigen Eigenschaften ausgewählt (4.5). Die Dokumentation der relevanten Spalten von `co6_config_variables` ist in der Tabelle 4.6 dargestellt.

Tabelle 4.6: Relevante Spalten von `co6_config_variables`.

Spalte	Datentyp	Information
id	bigint	Hauptschlüssel der Tabelle, z. B. 100
name	varchar	Name der Konfigurationsvariable in der Tabelle, z. B. AF
description	varchar	Beschreibung oder Name der Entitäten, z. B. Atemfrequenz
unit	varchar	Maßeinheit, z. B. 1/min
co6_config_variableTypes_id	varchar	Datentyp der gespeicherten Werte der Konfigurationsvariable, z. B. decimal_6_3
parent	varchar	Bezug der Konfigurationsvariable, z. B. Patient
deleted	varchar	null wenn die Konfigurationsvariable noch in Nutzung ist, sonst Datum der Löschung

Die Tabelle `co6_config_variable_types` enthält die Datentypen der Konfigurationsvariablen und, unter anderem, die Information in welchen Werttabellen die korrespondierenden Werte der Messungen oder Beobachtungen der Biosignale im COPRA sich befinden. Denn die Daten werden in COPRA in separaten Tabellen je nach Datentyp gespeichert. Die wichtigsten Spalten der Tabelle `co6_config_variable_types` sind in der Tabelle 4.7 aufgelistet.

Tabelle 4.7: Relevante Spalten von `co6_config_variable_types`.

Spalte	Datentyp	Information
id	bigint	Hauptschlüssel der Tabelle, z. B. 2
name	varchar	Name des Datentyps, z. B. String
tableName	varchar	Name der Tabelle, in der die Information im COPRA gespeichert wird, z. B. <code>co6_data_string</code>

Eine weitere wichtige Tabelle für die spätere Überführung der Daten ist `co6_medic_data_patient` (Tabelle 4.8). Diese beinhaltet die Basisinformation der behandelnden Personen. Diese Tabelle beinhaltet zusammen mit der pseudonymisierten Patientennummer einen Hauptschlüssel für die Verlinkung der Werte der Biosignaldaten in den Werttabellen mit der korrespondierenden Person. Aus diesem Grund befindet sich dieser Schlüssel als Fremdschlüssel auch in den Werttabellen.

Tabelle 4.8: Relevante Spalten von `co6_medic_data_patient`.

Spalte	Datentyp	Information
id	bigint	Hauptschlüssel der Tabelle, z. B. 25
patid	varchar	Patientennummer
deleted	boolean	false , wenn die Person noch im System ist, sonst true

4.4.1 Werttabellen

Die Werttabellen (Tabelle 4.9) beinhalten die gespeicherten Werte der Konfigurationsvariablen.

Diese Tabellen besitzen den Hauptschlüssel der Tabelle `co6_config_variables` als Fremdschlüssel, um die Werte der Biosignaldaten den korrespondierenden Konfigurationsvariablen zuzuordnen. Von diesen Tabellen, genau wie bei den anderen Tabellen, wurden nur die relevanten Spalten für die Überführung der Information von COPRA in FHIR berücksichtigt.

Tabelle 4.9: Werttabellen in der COPRA-Instanz des DWH. Alle Tabellen beinhalten die Hauptschlüssel der Tabellen `co6_medic_data_patient` und `co6_config_variables` als Fremdschlüssel. Die Tabellen `co6_data_decimal_6_3` und `co6_data_string` besitzen eine Spalte „`val`“ für die Speicherung der Werte oder Parameter der angewandten Techniken. Die Tabelle `co6_medic_pressure` wiederum beinhaltet drei Spalten für die Speicherung der systolischen, mittleren, und diastolischen Werte der Blutdruckmessungen.

Werttabelle	Datentypen
<code>co6_data_decimal_6_3</code>	numerische Werte
<code>co6_data_string</code>	Zeichenketten
<code>co6_medic_pressure</code>	Blutdruckmessungen

Die Werttabellen `co6_data_decimal_6_3` und `co6_data_string` besitzen eine ähnliche Struktur. Der Unterschied zwischen beiden Tabellen liegt darin, dass die Spalte `val` in der Tabelle `co6_data_decimal_6_3` die numerischen Werte der Messungen registriert und diese Spalte in der Tabelle `co6_data_string` die Zeichenketten-Elemente erfasst. Wie bei anderen Tabellen im COPRA-System wurden nur die wichtigen Attribute für diese Arbeit ausgewählt. Die Struktur der Tabellen `co6_data_decimal_6_3` und `co6_data_string` ist in der Tabelle 4.10 dargestellt.

Tabelle 4.10: Relevante Spalten von `co6_data_decimal_6_3` und `co6_data_string`.

Spalte	Datentyp	Information
<code>id</code>	<code>bigint</code>	Hauptschlüssel der Tabelle, z. B. 2595
<code>varid</code>	<code>int</code>	Fremdschlüssel, der zur Spalte <code>id</code> in der Tabelle <code>co6_config_variables</code> zeigt, z. B. 102
<code>deleted</code>	<code>boolean</code>	false , wenn der Wert noch im System ist, sonst true
<code>parent_id</code>	<code>bigint</code>	Fremdschlüssel, der zur Spalte <code>id</code> in der Tabelle <code>co6_data_patient</code> zeigt, wenn das Wert von <code>parent_id</code> gleich 1 ist
<code>parent_varid</code>	<code>int</code>	Fremdschlüssel, der zur Spalte <code>id</code> in der Tabelle <code>co6_config_variables</code> zeigt. Diese Spalte ist der Bezug der Biosignaldaten, z. B. 1 für Patientenbezug

datetimeto	timestamp	Datum und Uhrzeit der Messung in ISO 8601 Format (JJJJ-MM-DD HH:mm:ss)
validated	boolean	true wenn den Eintrag validiert wurde, sonst false .
flagcurrent	boolean	true wenn den Eintrag noch aktuell ist, sonst false .
val	numeric/varchar	Wert der Biosignaldaten. Der Wert ist eine Zahl in der Tabelle co6_data_decimal_6_3, z. B. 38, und eine Zeichenkette in der Tabelle co6_data_string, z. B. blass

Die Werttabelle **co6_medic_pressure** (Tabelle 4.11) beinhaltet die systolischen, mittleren und diastolischen Blutdruckwerte.

Tabelle 4.11: Relevante Spalten von co6_medic_pressure.

Spalte	Datentyp	Information
id	bigint	Hauptschlüssel der Tabelle, z. B. 2595
varid	int	Fremdschlüssel, der zur Spalte id in co6_config_variables zeigt, z. B. 102
deleted	boolean	false , wenn der Wert noch im System ist, sonst true
parent_id	bigint	Fremdschlüssel, der zur Spalte id in der Tabelle co6_data_patient zeigt, wenn das Wert von parent_id gleich 1 ist
parent_varid	int	Fremdschlüssel, der zur Spalte id in co6_config_variables zeigt. Diese Spalte ist der Bezug der Biosignaldaten
datetimeto	timestamp	Datum und Uhrzeit der Messung in ISO 8601 Format (JJJJ-MM-DD HH:mm:ss)
validated	boolean	true wenn den Eintrag validiert wurde, sonst false .
flagcurrent	boolean	true wenn den Eintrag aktuell ist, sonst false .
systolic	numeric	Systolischer Blutdruckwert, z. B. 120
mean	numeric	Mittel Blutdruckwert, z. B. 100
diastolic	numeric	Diastolischer Blutdruckwert, z. B. 80

Die Zusammenführung von Parametern der Tabellen `co6_config_variables`, `co6_medic_data_patient` und der Werttabellen, zusammen mit manche Komponenten der Tabelle `mii_icu` bilden am Ende die FHIR-Ressourcen des Erweiterungsmoduls „Intensivmedizin“ ab.

4.5 Auswahl der Konfigurationsvariablen von COPRA

Die Tabelle `co6_config_variables` in Zusammenhang mit den Werttabellen wurde untersucht, um die relevanten Konfigurationsvariablen für die Durchführung der Arbeit zu erkennen. Die Tabelle `co6_config_variables` beinhaltet 7409 Konfigurationsvariablen. Die Kriterien für die Wahl der Variablen sind Folgende:

- Variablen in Benutzung: Spalte `deleted` der Tabelle der Konfigurationsvariablen hat den Wert `null`
- Variablen mit validierten Werten in den Werttabellen: Spalte `validated` hat den Wert `true`
- Variablen mit nicht gelöschten Werten in den Werttabellen: Spalte `deleted` der Werttabelle hat den Wert `false`
- Variablen mit aktuellen Werten in den Werttabellen: Spalte `flagCurrent` der Werttabelle hat den Wert `true`
- Variablen mit Patienten- oder Fallbezug: Spalte `parent` der Tabelle der Konfigurationsvariablen hat die Werte 1 für Patientenbezug oder 20 für Fallbezug
- Variablen mit 1000 Werten oder mehr in den Werttabellen

Um die Konfigurationsvariablen auszuwählen, wurde für jede Werttabelle eine SQL-Abfrage, wie im Code 4.1, realisiert.

```
select
count(cvt.id) quantity_in_value_table, -- Menge an Werten je
Konfigurationsvariable in der Werttabelle
ccv.id, -- ID der Konfigurationsvariable
ccv."name" -- Name der Konfigurationsvariable
from copra.co6_config_variables ccv -- Tabelle der
Konfigurationsvariablen
join copra.value_table cvt -- eine der Werttabellen
```

```

on cvt.varid = ccv.id -- Verlinkung der Werttabelle mit der Tabelle
  co6_config_variables
where not cdd.deleted -- nicht gelöschte Daten
and ccv.parent in (1, 20) -- Patienten- oder Fallbezug
and not ccv.deleted -- Konfigurationsvariable wird benutzt
and cvt.flagcurrent -- aktuelle Daten
and cvt.validated -- validierte Daten
group by ccv.id, ccv."name"
having count(cvt.id) >= 1000 -- 1000 Werte oder mehr in den Werttabellen
order by quantity_in_value_table -- Menge an Werten absteigend sortiert
;

```

Listing 4.1: SQL-Abfrage zur Auswahl der Konfigurationsvariablen im COPRA-System. „copra.value_table cvt“ stellt die verschiedenen Werttabellen im System dar: co6_data_decimal_6_3, co6_data_string und co6_medic_pressure. „quantity_in_value_table“ ist ein Alias für die Spalte der Menge an Werten in den Werttabellen.

Von den 7409 Konfigurationsvariablen in der COPRA-Instanz wurden 701 als relevante Variablen identifiziert. Unter diesen befinden sich die Biosignalparameter für die Zuordnung mit den FHIR-Profilen.

Die Anzahl der ausgewählten Konfigurationsvariablen in den Werttabellen für das Pattern Matching der Variablen mit den FHIR-Profilen für die Erkennung der Biosignaldaten in COPRA ist in der Tabelle 4.12 dargestellt.

Tabelle 4.12: Anzahl der repräsentierten Konfigurationsvariablen in je Werttabelle. Die meisten Konfigurationsvariablen sammeln numerische Werte.

Anzahl	Werttabelle
492	co6_data_decimal_6_3
199	co6_data_string
10	co6_medic_pressure
701	Gesamt

Unter der 701 ausgewählten Konfigurationsvariablen befinden sich die Biosignaldaten für die Datenzuordnung mit den FHIR-Profilen.

4.6 Analyse der Tabellen für die Datendefinition

Mit der Information der relevanten Parameter von COPRA zusammen mit den Parametern der FHIR-Profile, wurden die geeigneten Spalten der Tabellen von COPRA und der Tabelle mii_icu in der DB mii_copra ausgewählt,

die die Parameter für die Überführung der Biosignaldaten in FHIR beinhalten.

Die Spalten **profile_name** und **loinc** in der Tabelle **mii_icu** (Tabelle 4.1), und die Spalten **name** und **description** in der Tabelle **co6_config_variables** (Tabelle 4.6) wurden als geeignete Kandidaten ausgewählt, um die Konfigurationsvariablen aus COPRA mit den FHIR-Profilen miteinander zu verlinken, denn diese Spalten beinhalten die Hauptinformation für den späteren Zuordnungsprozess.

Einerseits beinhaltet die Spalte **profile_name** die Namen der FHIR-Profile, die wiederum auch die Namen von Verfahren oder Messungen sind, andererseits registriert die Spalte **loinc** die LOINC-Codes. Der Vorteil dieses Codesystem anzuwenden, liegt daran, dass LOINC das meist verwendete Codesystem in den FHIR-Profilen (Tabelle 4.3) ist.

Nach der Analyse der gespeicherten LOINC-Codes wurden Abkürzungen gefunden, die für den späteren Pattern Matching-Prozess von Bedeutung sind, denn sie dienen als zusätzliche Namen „Short Name“ in der Beschreibung der LOINC-Codes. Manche dieser „Short Names“ werden auch in der alltäglichen Kommunikation unter dem Gesundheitspersonal benutzt, wie die Abkürzung FIO2 für die Sauerstofffraktion (Tabelle 4.13). Solche „Short Names“ mit einigen Variationen wurden auch in den Namen und Beschreibungen von den Konfigurationsvariablen in COPRA durch das Pattern Matching erkannt.

Tabelle 4.13: Beispiel von LOINC-„Short Name“. Das FHIR-Profil „Sauerstofffraktion“ hat den LOINC-Code, 71835-3. Dieser wiederum besitzt den „Short Name“ FIO2.

Profilname	LOINC-URL	Short Name
Sauerstofffraktion	https://loinc.org/71835-3/	FIO2

In der Tabelle **co6_config_variables** sind die an dem Standort angegebenen COPRA-Namen für Messungen oder Techniken. Diese sind in der Spalte **name** registriert.

Die COPRA-Namen können Fachtermini oder eine Zusammensetzung von Namensabkürzungen des angewandten Verfahrens und Gerätes sein. Manche der Abkürzungen des Verfahrens sind Varianten des vorher erläuterten LOINC-„Short Name“. Die Spalte **description** in **co6_config_variables** beinhaltet eine kurze Beschreibung der Verfahren und in manchen Fällen befindet sich in dieser Beschreibung der ganze Name eines Verfahrens. Diese Details führten dazu, dass die Spalte **description** für die Zuordnung der Konfigurationsvariablen mit den FHIR-Profilen betrachtet werden sollte.

Die Maßeinheiten der Konfigurationsvariablen und somit der Biosignaldaten befinden sich in der Spalte `unit` der Tabelle `co6_config_variables`. Andererseits befinden sich die Maßeinheiten der Profile in der Spalte `valuequantity_code` der Tabelle `mii_icu`. Die Maßeinheiten sind entscheidend, da diese in beiden Systemen für die Harmonisierung vergleichbar sein müssen. Das bedeutet, dass die Maßeinheiten beider Systeme dieselben Dimensionen der physikalischen Größen besitzen müssen, sodass eine Umrechnung zwischen den Einheiten beider Systeme möglich ist.

Dadurch, dass ein Identifikator der Patienten in den FHIR-Profilen als Referenz zu der FHIR-Ressource „Patient“ im Kerndatensatz der MII benutzt wird, um die Biosignaldaten zu den korrespondierenden, behandelnden Personen zu verlinken, wird dieser Parameter vom COPRA-System benötigt. Diese Information befindet sich in der Tabelle `co6_medic_patient` (Tabelle 4.8) in der Spalte `patid`. Der Inhalt dieser Spalte ist ein Pseudonym der Patientennummer im Krankenhausinformationssystem (KIS). Auch in der Tabelle `co6_medic_patient` soll die Spalte `id` berücksichtigt werden, denn diese ist der primäre Schlüssel dieser Tabelle und ist somit Fremdschlüssel in den Werttabellen. Damit werden die Daten der Patienten mit ihren Biosignaldaten im COPRA-System verknüpft. Die Spalte `deleted` in der Tabelle `co6_medic_patient` wird benötigt, um über die Löschung einer Person aus dem System in Kenntnis gesetzt zu werden. Die Abbildung von `patid` in den FHIR-Profilen ist das Attribut `reference` des Elements `subject`.

Die Werte und Zeitangaben der Biosignale befinden sich in den Werttabellen. Diese Werte sind in der Spalte `val` der Tabellen `co6_data_decimal_6.3` und `co6_data_string` (Tabelle 4.10), und in den Spalten `systolic`, `mean` und `diastolic` der Tabelle `co6_medic_pressure` (Tabelle 4.11) zu finden. Die Zeiten der Messungen sind in der Spalte `datetimeto` aller Werttabellen zu finden. Die Abbildung der Werte und Zeitangaben der Biosignale in den FHIR-Profilen sind das Attribut `value` des Elements `valueQuantity` beziehungsweise das Element `effectiveDateTime`.

Die Werttabellen beinhalten den Hauptschlüssel der Tabelle `co6_medic_patient` in der Spalte `parent_id` als Fremdschlüssel, wenn der Wert dieser Spalte gleich 1 ist. Damit wird die Information der Biosignaldaten den behandelnden Personen zugewiesen. Die Spalte `var_id` ist der Schlüssel, welcher die Werte den korrespondierenden Konfigurationsvariablen in der Tabelle `co6_config_variables` zuweist. Der Hauptschlüssel einer Werttabelle in der Spalte `id` in Kombination mit dem Namen der Werttabelle wird benutzt, um einen Ersatzschlüssel für die Überführung der Daten von COPRA in FHIR zu erstellen.

Nach der Datendefinition wird die Zuordnung der Konfigurationsvariablen mit den FHIR-Profilen durch das Pattern Matching durchgeführt. Das

Ergebnis dieses Prozesses wird in der Tabelle `mapping_mii_co6` (Tabelle 4.14) gespeichert.

Tabelle 4.14: Tabelle für die Speicherung der Zuordnung der Konfigurationsvariablen mit den FHIR-Profilen nach dem Pattern Matching. Die Spalten mit `coding_system` in den Namen speichern die Uniform Resource Identifier (URI) (Anhang B) der Terminologie-Systeme.

Spalte	Datentyp	Information
<code>profile_id</code>	int	Generierter numerischer Identifikator des zugeordneten FHIR-Profiles
<code>profile_name</code>	varchar	Name des FHIR-Profiles, z. B. Atemzugvolumen-Während-Beatmung
<code>category_coding_system</code>	text	URL der Kategorie. Dies kann ein SNOMED CT-URL bei einigen „Observations“ sein.
<code>category_coding_code</code>	varchar	„vital-signs“ oder SNOMED CT-ID
<code>code_coding_system_snomed</code>	text	http://snomed.info/sct
<code>code_coding_code_snomed</code>	varchar	SNOMED CT des FHIR-Profiles, z. B. 250874002
<code>code_coding_system_loinc</code>	text	http://loinc.org
<code>code_coding_code_loinc</code>	varchar	LOINC-Code des FHIR-Profiles, z. B. 76222-9
<code>code_coding_system_ieee</code>	text	urn:iso:std:iso:11073:10101
<code>code_coding_code_ieee</code>	varchar	ISO/IEEE 11073-10101™-Schlüssel des FHIR-Profiles, z. B. 151980
<code>valuequantity_system</code>	text	http://unitsofmeasure.org
<code>valuequantity_code</code>	varchar	Maßeinheiten im FHIR-Profil, z. B. mL
<code>conf_var_unit</code>	varchar	Maßeinheiten der Konfigurationsvariable in COPRA, z. B. ml

device.reference	text	Zuweisung zur Art der Prozedur (gemessen, eingestellt oder erhoben). Diese Prozedur befindet sich in den Profilen des Typs „DeviceMetric“
meta_profile	text	URL um das Profil zu identifizieren
conf_var_id	bigint	Primärer Schlüssel der Tabelle <code>co6_config_variables</code> , z. B. 104726
conf_var_parent_id	bigint	Schlüssel des Bezuges der Variable, z. B. 1 für Patientenbezug
conf_var_parent_name	varchar	Name des Bezuges, z. B. Patient
conf_var_name	varchar	Name der Konfigurationsvariable, z. B. Beatmung_MS_Pallas_Vt
conf_var_description	varchar	Beschreibung der Konfigurationsvariable, z. B. gemessenes Tidalvolumen
conf_var_types_id	int	Schlüssel des Datentyps in der Tabelle <code>co6_config_variable_types</code> , z. B. 6 für numerische Werte
conf_var_types_name	varchar	Name des Datentyps in der Tabelle <code>co6_config_variable_types</code> , z. B. Decimal_6_3
code_systolic_coding_system_snomed	text	http://snomed.info/sct
code_systolic_coding_code_snomed	varchar	SNOMED CT-ID der systolischen Blutdruckmessung, z. B. 271649006
code_systolic_coding_system_loinc	text	http://loinc.org
code_systolic_coding_code_loinc	varchar	LOINC-Code der systolischen Blutdruckmessung, z. B. 8406-1
code_systolic_coding_system_ieee	text	urn:iso:std:iso:11073:10101

code_systolic _coding_code_ieee	varchar	ISO/IEEE 11073-10101™-Schlüssel der systolischen Blutdruckmessung, z. B. 150107
code_mean _coding_system _snomed	text	http://snomed.info/sct
code_mean _coding_code _snomed	varchar	SNOMED CT-ID der mittleren Blutdruckmessung, z. B. 6797001
code_mean _coding_system _loinc	text	http://loinc.org
code_mean _coding_code_loinc	varchar	LOINC-Code der mittleren Blut- druckmessung, z. B. 8478-0
code_mean _coding_system _ieee	text	urn:iso:std:iso:11073:10101
code_mean _coding_code_ieee	varchar	ISO/IEEE 11073-10101™-Schlüssel der mittleren Blutdruckmessung, z. B. 150019
code_diastolic _coding_system _snomed	text	http://snomed.info/sct
code_diastolic _coding_code _snomed	varchar	SNOMED CT-ID der diastolischen Blutdruckmessung, z. B. 271650006
code_diastolic _coding_system _loinc	text	http://loinc.org
code_diastolic _coding_code_loinc	varchar	LOINC-Code der diastolischen Blut- druckmessung, z. B. 8462-4
code_diastolic _coding_system _ieee	text	urn:iso:std:iso:11073:10101
code_diastolic _coding_code_ieee	varchar	ISO/IEEE 11073-10101™-Schlüssel der diastolischen Blutdruckmessung, z. B. 150018
matching_valide	bool	Ergebnis der Validierung der Zuord- nung, true oder false

4.7 Pattern Matching der FHIR-Profile mit den Konfigurationsvariablen

Nach der Erkennung der 701 relevanten Konfigurationsvariablen und der Datendefinition, wurde das Pattern Matching durchgeführt, um bestimmte Muster zwischen Parametern der FHIR-Profile und Attributen der Konfigurationsvariablen zu erkennen.

Zu Beginn wurde eine unscharfe Suche durchgeführt, um die FHIR-Profile den Konfigurationsvariablen zuzuordnen. Für diesen Zweck wurden der Parameter `profile_name` der Tabelle `mii_icu` zusammen mit den Parametern `name` und `description` der Tabelle `co6_config_variables` in einer Comma-Separated Values (CSV)-Datei extrahiert.

Mit einem Python-Skript (Code 4.2) und der CSV-Datei wurde die unscharfe Suche durchgeführt. Dieser Prozess liefert einen Datensatz mit 400 Paaren von zugeordneten Profilen mit Konfigurationsvariablen. 355 Paare davon waren falsche Zuordnungen, wie die Beispiele der Tabelle 4.15 zeigen.

```
from thefuzz import process, fuzz # Bibliotheken der unscharfe Suche
    laden

with open("profils_names.csv") as fp: # Datei mit den Namen der Profile
    prn = fp.readlines() # Name der Profile
    prn = [x.strip() for x in prn] # Formatierung

with open("config_vars_names.csv") as fc: # Datei mit den
    Konfigurationsvariablen
    cvn = fc.readlines() # Konfigurationsvariablen
    cvn = [x.strip() for x in cvn] # Formatierung

for profil in prn: # Durchlauf die Namen der Profile
    match_ratios = process.extract(profil, cvn, scorer=fuzz.token_set_ratio)
    # Liste der passenden Konfigurationsvariablen.
    for prn_match_cvn in match_ratios: # Durchlauf die
        Konfigurationsvariablen
        print(profil, prn_match_cvn) # Liste der Paare "Profile -
        Konfigurationsvariablen"
```

Listing 4.2: Python-Skript für das Pattern Matching. Die Option „`token_set_ratio`“ ermöglicht eine flexiblere fuzzy-Suche (2.8.1), denn die Länge der Namen der Profile und die Länge der Namen und Beschreibungen der Konfigurationsvariablen sind deutlich unterschiedlich.

Tabelle 4.15: Beispiele von falsch zugeordneten FHIR-Profilen. Die Namen der Konfigurationsvariablen und deren Beschreibung sind durch „-“ getrennt.

Profile	Konfigurationsvariablen
Arterieller Druck	ICP - Intrakranialer Druck
Atemfrequenz	HF - Herzfrequenz
Rechtsatrialer Druck	CPP - Zerebraler Perfusionsdruck
Venoeser Druck	PAP - Pulmunalarterieller Druck

Nach dem vorherigen Ergebnis der unscharfen Suche wurden die FHIR-Profile und die Konfigurationsvariablen nochmal analysiert. Dieses Mal mit SQL-Abfragen mit Regexs. Für dieses Verfahren wurden die Informationen der Spalten `profile_name` und `loinc` der Tabelle `mii_icu` benutzt, um Muster in den Spalten `name` und `description` der Tabelle `co6_config_variables` zu identifizieren. Die Abbildung 4.3 zeigt eine Darstellung der angewandten Parameter der Tabellen und deren Zusammenführung im Pattern Matching-Prozess.

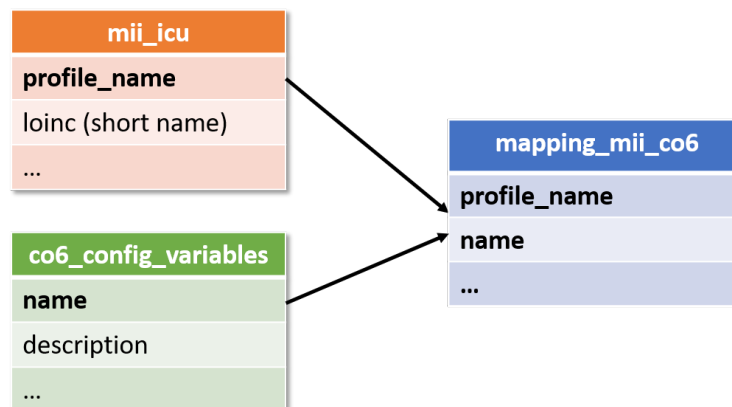


Abbildung 4.3: Schematische Darstellung der Zuordnung der Spalte `profile_name` und die „Short Names“ der LOINC-Codes der Spalte `loinc` der Tabelle `mii_icu` mit den Spalten `name` und `description` der Tabelle `co6_config_variables`. Das Ergebnis der Zuordnung wird in der Tabelle `mapping_mii_co6` gespeichert.

Wichtig für die Definition der Regexs ist die Auswahl der Teilwörtern (2.8) oder Kombinationen davon, die die FHIR-Profile charakterisieren. Denn diese Teilwörter sind in den meisten Fällen Fachtermini oder Teile von Fachbegriffen und können wieder in den Namen oder Beschreibungen der Konfigurationsvariablen gefunden werden.

Manche Teilwörter in den Profilnamen sind sehr allgemein und befinden sich in mehreren Profilen und in zahlreichen Namen und Beschreibungen der Konfigurationsvariablen, z. B. „druck“. Das hat zur Folge, dass viele Konfigurationsvariablen falsch einem Profil durch die Nutzung von nur einem von diesen Teilwörtern zugeordnet werden könnten. Solche Teilwörter wurden in Kombination mit anderen benutzt, um das Pattern Matching zu verfeinern und die Trefferquote zu erhöhen. Hier ist auch die Menge und Länge der angewandten Teilwörter entscheidend, denn die Anwendung von vielen oder zu langen Teilwörtern liefert auch kein Ergebnis. Aus diesem Grund wurde bei der Schreibung der SQL-Abfragen je FHIR-Profil solange die Definition des Regex getestet, bis die optimale Kombination von Menge und Länge der Teilwörter gefunden wurde, die die genaue Menge an Konfigurationsvariablen geliefert hatte.

Der Code 4.3 zeigt beispielhaft die SQL-Abfrage mit Regex für die gleichzeitige Erkennung mehrerer allgemeiner Teilwörter des Profils „Mittlerer Beatmungsdruck“ in der Tabelle der Konfigurationsvariablen, um das FHIR-Profil den Konfigurationsvariablen zuzuordnen.

```
select
  name,
  description,
  id,
  unit
from mii_copra.co6_config_variables
where name ~* 'mitt.+atm.+druc'
or description ~* 'mitt.+atm.+druc'
or name ~* 'mea.+pres.+(resp|vent)'
or description ~* 'mea.+pres.+(resp|vent)';
```

Listing 4.3: SQL-Abfrage mit allgemeinen Teilwörtern.

Im Code 4.3 werden die Spalten `name` und `description` der Tabelle `co6_config_variables` gleichzeitig zur Erkennung von Mustern untersucht. In der definierten Regex dieser SQL-Abfrage wird die Tilde „~“ gefolgt von einem Sternchen „*“ benutzt, um Groß- und Kleinschreibung nicht zu berücksichtigen. Die Teilwörter „mitt“, „atm“ und „druc“ des Profils „Mittlerer Beatmungsdruck“ verbunden durch „+“, für eine Folge von beliebigen Zeichen, haben das Ergebnis der Tabelle 4.16 geliefert. Die Teilwörter „mea“, „pres“ und die Gruppierung mit den Alternativen „resp“ oder „vent“ wurden benutzt, um den Name (Mean pressure Respiratory system airway –on ventilator) auf Englisch und den „Short Name“ (Mean Pres on vent Airway) zu berücksichtigen.

Tabelle 4.16: Erkannte Konfigurationsvariablen bei der Anwendung der Kombination von allgemeinen Teilwörtern. Das FHIR-Profil „Mittlerer Beatmungsdruck“ wurde drei Konfigurationsvariablen zugeordnet.

COPRA-Name	COPRA-Beschreibung
Beatmung_MS_VisionA_MAP	gemessener Mittlerer Atemwegsdruck unter HFO bei Alpha Vision
Beatmung_ES_VisionA_MAP	Mittlerer Atemwegsdruck (MAP)
Beatmung_MS_Servoi_Pmean	Mittlerer Atemwegsdruck

Andere Teilwörter wie „intrakraniell“ können hingegen alleine benutzt oder zerteilt werden, ohne die Trefferquote des Pattern Matchings zu reduzieren, weil sie nicht sehr häufig in den Profilnamen und in den Namen und Beschreibungen der Konfigurationsvariablen vorkommen. Solche Teilwörter charakterisieren besser ein FHIR-Profil. Der Code 4.4 zeigt beispielhaft die SQL-Abfrage mit der Definition des Regex mit nur einem Teilwort, um die passenden Konfigurationsvariablen zu dem FHIR-Profil „Intrakranieller Druck“ zu detektieren.

```
select
  name,
  description,
  id,
  unit
from mii_copra.co6_config_variables
where name ~* 'intra[ck]'
or description ~* 'intra[ck]'
;
```

Listing 4.4: SQL-Abfrage mit einem seltenen Teilwort.

Im Code 4.4 werden dieselben Spalten, wie beim Code 4.3 gleichzeitig untersucht. In diesem Beispiel befinden sich in den eckigen Klammern die Buchstaben „c“ und „k“, um die Teilwörter „intrakraniell“, oder „intracranial“ auf Englisch, zu identifizieren.

Ein vorher genannter Aspekt ist die allgemeine, umgangssprachliche Anwendung der LOINC-„Short Names“ vom Personal des Gesundheitswesens und wie dieser Parameter in der Konfigurationsvariablen zu finden ist. Ein Beispiel des Ergebnisses eines Pattern Matchings, um „Short Names“ der LOINC-Codes zu berücksichtigen, ist in der Tabelle 4.17 dargestellt. In diesem Beispiel ist die Erkennung eines „Short Name“ in der Spalte **name** der Tabelle **co6_config_variables** repräsentiert.

Tabelle 4.17: Erkennung von „Short Names“ von Verfahren in der Tabelle `co6_config_variables`. Der „Short Name“ des LOINC-Codes 8837-7 vom FHIR-Profil „Systemischer Vaskulaerer Widerstandsindex“ ist „SV RI“. Diese Abkürzung, ohne Leerzeichen, wird somit in der Spalte `name` der Tabelle `co6_config_variables` wiedererkannt, nämlich SVRI.

Profilname	loinc	Short Name	COPRA-Name
Systemischer Vaskulaerer Widerstandsindex	8837-7	SV RI	SVRI

Am Ende des Pattern Matching-Prozesses konnten von den 701 Konfigurationsvariablen 75 einem FHIR-Profil zugeordnet werden, und von den 80 FHIR-Profilen wurden dann 41 mit einer COPRA-Variable verlinkt. Das sind insgesamt 80 importierte Einträge in der Tabelle `mapping_mii_co6`, weil ein Profil mehreren Konfigurationsvariablen zugeordnet wurde. Alle zugeordneten Konfigurationsvariablen waren patientenbezogene Biosignalparameter in COPRA.

Das Ergebnis des Pattern Matchings ist in der Tabelle `mapping_mii_co6` (Tabelle 4.14) registriert. Dieser Datensatz wurde danach (4.8) von den Fachleute der PDMS-Abteilung validiert.

4.8 Validierung des Pattern Matchings

Nach dem Pattern Matching wurde ein Meeting mit dem Fachpersonal der PDMS-Abteilung organisiert, um den resultierenden Datensatz der Zuordnung zu analysieren und zu validieren. Dieser Datensatz wurde vorab an die fachlichen Spezialisten gesendet.

Im generierten Datensatz des Pattern Matchings wurden 18 Konfigurationsvariablen ohne Maßeinheiten erkannt. Eine Liste dieser Konfigurationsvariablen wurde auch den Spezialisten der PDMS-Abteilung zur Verfügung gestellt. Die Fachleute der PDMS-Abteilung konnten mit Hilfe von dieser Liste im Frontend des COPRA-Systems nach den fehlenden Einheiten suchen.

Am Ende wurden die Maßeinheiten von 16 Konfigurationsvariablen im COPRA identifiziert und nachträglich in die Spalte `conf_var_unit` der Tabelle `mapping_mii_co6` eingefügt.

Die Analyse des Datensatzes des Pattern Matchings zur Validierung führte zum Ergebnis, dass die 75 zugewiesene Konfigurationsvariablen richtig den jeweiligen FHIR-Profilen zugeordnet wurden.

4.9 Harmonisierung der Maßeinheiten

Nach der Validierung des Ergebnisses des Pattern Matchings und mit den eingefügten Maßeinheiten in die Tabelle der resultierenden Zuordnung, folgte eine detaillierte Untersuchung der Maßeinheiten in COPRA und einen Vergleich zwischen den Maßeinheiten beider Systeme für die Harmonisierung der Einheiten beider Systeme.

Die zwei Paare von zugeordneten Profilen mit Konfigurationsvariablen bei denen die Konfigurationsvariablen keine Maßeinheiten weder in der Tabelle `co6_config_variables` noch im Frontend von COPRA erhalten haben (Tabelle 4.18), werden in diesem Projekt nicht weiter betrachtet, weil solche Werte mit unbekannten Maßeinheiten in COPRA sich nicht harmonisieren lassen.

Tabelle 4.18: Konfigurationsvariablen ohne Maßeinheiten.

Profilname	COPRA-Name	Profil-Einheit	COPRA-Einheit
Ionisiertes Kalzium aus Nierenersatzverfahren	NEV_CRRT_VO_CalciumLoesung	mmol/L	NULL
Zeitverhaeltnis-Ein-Ausatmung	Beatmung_MS_G5_IE_Verhaeltnis	{ratio}	NULL

Nach der Untersuchung der Maßeinheiten in COPRA wurden 38 unterschiedliche Konfigurationsvariablen mit derselben Maßeinheit wie bei den zugewiesenen FHIR-Profilen, aber mit unterschiedlichen Schreibweisen, erkannt. Die Tabelle 4.20 zeigt einige Beispiele davon.

Tabelle 4.19: Schreibweise derselben Maßeinheiten in beiden Systemen.

Profil-Einheit	COPRA-Einheit	Anmerkung
cm[H2O]	cmH2O	-
Cel	C°	-
{Breaths}/min	AZ/min	Atemzüge pro Minute

Auch dieselbe Maßeinheit wurde zwischen Konfigurationsvariablen unterschiedlich dargestellt. Dieses Phänomen wurde in sieben Fälle gefunden. Die Tabelle 4.20 zeigt einige Beispiele dieses Ereignisses.

Tabelle 4.20: Beispiel der Schreibweise derselbe Maßeinheit bei unterschiedlichen Konfigurationsvariablen in COPRA. bpm steht für „breaths per minute“.

Profilname	COPRA-Name	Profil-Einheit	COPRA-Einheit
Mittlerer Beatmungsdruck	Beatmung_MS_VisionA_MAP	cm[H2O]	cm H2O
Mittlerer Beatmungsdruck	Beatmung_MS_Servoi_Pmean	cm[H2O]	[cmH2O]
Mechanische-Atemfrequenz-Beatmet	Beatmung_MS_Evita4_frequenz	{Breaths}/min	bpm
Mechanische-Atemfrequenz-Beatmet	Beatmung_MS_G5_ftotal	{Breaths}/min	AZ/min

In der Untersuchung der Maßeinheiten wurde auch eine fehlerhafte Darstellung der Maßeinheit bei einer Konfigurationsvariable festgestellt. Diese Problematik konnte auch nicht von den Experten der PDMS-Abteilung geklärt werden. Aus diesem Grund wird diese Konfigurationsvariable nicht weiter betrachtet. Die Tabelle 4.21 zeigt die Konfigurationsvariable mit diesem Problem.

Tabelle 4.21: Fehlerhafte Darstellung Maßeinheiten in COPRA.

Profilname	COPRA-Name	Profil-Einheit	COPRA-Einheit
Linksventrikulärer Schlagvolumenindex	Vigileo_SVI	mL/m2	ml/b/m ² ml/b/m ²

Es wurden acht Konfigurationsvariablen mit Maßeinheiten identifiziert, die nicht dieselben Maßeinheiten haben wie die Einheiten der FHIR-Profile. Die Maßeinheiten von sieben dieser Konfigurationsvariablen konnten in die Maßeinheiten der FHIR-Profile umgerechnet werden. Diese Umwandlung wird durch die Multiplikation der numerischen Werte der Werttabellen mit Umrechnungsfaktoren realisiert. Die Tabelle 4.22 zeigt die Liste der betroffenen Maßeinheiten.

Tabelle 4.22: Maßeinheiten und Umrechnungsfaktoren für die Harmonisierung. Die 1 in den Einheiten der FHIR-Profile bedeutet „decimal fraction“ [67]. Die Werte der Konfigurationsvariablen mit den gleichen Maßeinheiten wie die FHIR-Profile werden mit 1 multipliziert.

COPRA-Einheit	Profil-Einheit	Faktor/Konversion
mbar	cm[H2O]	1.01972
%	1	0.01
mmHg	cm[H2O]	1.35951
min	h	0.016667
COPRA-Einheit = Profil-Einheit		1

Andere Maßeinheiten von Konfigurationsvariablen können nicht in die Einheiten der FHIR-Profile umgewandelt werden, denn die Dimensionen der physikalischen Größen der Maßeinheiten sind in beiden Systemen nicht dieselben. Diese Konfigurationsvariablen können nicht harmonisiert werden. Aus diesem Grund werden sie in diesem Projekt nicht weiter betrachtet. Die Tabelle 4.23 zeigt die betroffenen Maßeinheiten.

Tabelle 4.23: Inkompatible Maßeinheiten.

Profilname	COPRA-Name	COPRA-Description	Profil-Einheit	COPRA-Einheit
Linksventrikulärer Herzindex	dPmax	Index der linken Ventrikelkontraktilität	L/(min.m2)	mmHg/s

Die Inkompatibilität dieser Einheiten liegt daran, dass die Dimensionen von L/(min.m2) Volumen, Zeit und Flächeninhalt sind. Die Dimensionen von mmHg/s sind wiederum Druck und Zeit. mmHg/s wird für das Ablassen der Manschette bei der Blutdruckmessung benutzt. In COPRA hingegen wird diese Maßeinheit an einer Konfigurationsvariable für die Messung des Indexes der linken Ventrikelkontraktilität gesetzt.

Am Ende dieser Analyse bleiben 76 Einträge im Datensatz der zugeordneten Paare der Konfigurationsvariablen mit den FHIR-Profilen mit insgesamt 71 Konfigurationsvariablen und 39 FHIR-Profilen. Der resultierende Datensatz mit den Umwandlungen für die Harmonisierung der Maßeinheiten beider Systeme wurde in die Tabelle `mapping_mii_co6_2` importiert. Diese neue Tabelle hat wiederum dieselbe Struktur wie `mapping_mii_co6` (Tabelle 4.14) mit einer neuen Spalte `unit_transform` des Datentyps „numeric“ mit dem Faktor für die Umrechnung der gespeicherten Biosignalen in den Maßeinheiten der FHIR-Profilen. Somit sind die Einheiten beider Systeme harmonisiert.

4.10 Zuweisung der Quell- und Zielfelder und weitere Transformationsregeln

Für die Zuweisung der Quell- und Zielfelder und die Definition von weiteren Transformationsregeln im Data Mapping sollen die Werte der Biosignaldaten von COPRA und deren Attribute mit den Parametern der FHIR-Profile zusammengeführt werden.

Die Werte der Biosignaldaten für die Erzeugung der FHIR-Ressourcen des Erweiterungsmoduls „Intensivmedizin“ befinden sich in der COPRA-Instanz des Staging Bereichs des DWH des DIZ der Universitätsmedizin Mainz. Die Datensätze für die Überführung der Information liegen in mehreren Tabellen und müssen hierzu im Regelfall zusammengeführt werden. Diese Tabellen wurden vorher erkannt und analysiert (4.6). Die Tabellen in COPRA und die dazu für die Überführung der Daten notwendigen Attribute sind in der folgenden Liste zusammengefasst:

- **co6_config_variables**: Namen und Schlüssel der Konfigurationsvariablen, die den FHIR-Profilen zugeordnet wurden
- **co6_medic_patient**: Pseudonymisierte Patientennummer und interne Identifikatoren der behandelnden Personen
- **co6_data_decimal_6_3**: Numerische Werte der Biosignale, Datum und Uhrzeit der Messung, interner Identifikator der Patienten und Schlüssel der Konfigurationsvariablen
- **co6_data_string**: String-Werte der Biosignale, Datum und Uhrzeit der Messung, interner Identifikator der Patienten und Schlüssel der Konfigurationsvariablen
- **co6_medic_pressure**: Systolische, mittlere und diastolische Blutdruckwerte, Datum und Uhrzeit der Messung, interner Identifikator der behandelnden Personen und Schlüssel der Konfigurationsvariablen

Die FHIR-Attribute für die Überführung der Biosignaldaten zusammen mit den Faktoren für die Umrechnung der COPRA-Maßeinheiten befinden sich in der Tabelle **mapping_mii_co6.2** (4.9) der DB **mii_copra**. Für die Arbeit mit den Daten dieser Tabelle in der COPRA-Instanz des Staging Bereichs des DWH wurde eine gleichnamige Tabelle in der COPRA-Instanz in SQL codiert. Danach wurde der Inhalt von **mapping_mii_co6.2** in der DB in einer CSV-Datei kopiert und in die erzeugte Tabelle **mapping_mii_co6.2**

der COPRA-Instanz importiert. Diese Tabelle im DWH dient der Zusammenführung der Werte der Biosignaldaten in COPRA mit den Elementen der FHIR-Profile.

Die verlinkten FHIR-Profile zu den Konfigurationsvariablen mit numerischen Werten und String-Werten sind ähnlich aufgebaut. Im Gegensatz dazu haben die Profile für Blutdruckmessungen eine etwas andere Struktur, trotzdem sind die Blutdruckmessungen-Profile untereinander auch gleich strukturiert (2.2.1). Diese strukturelle Ähnlichkeit der FHIR-Profile erleichtert die Modellierung der Zuweisung der Quell- und Zielfelder und die Definition von weiteren Transformationsregeln (4.10.1).

Für die Zuweisung der Quell- und Zielfelder werden weitere Transformationen definiert. Eine davon ist die Nutzung der, im PostgreSQL implementierten, MD5-Hashfunktion zur Erzeugung der eindeutigen Identifikatoren für die FHIR-Ressourcen.

Eine weitere Transformation ist die Umwandlung des Datentyps „type casting“ bei manchen Werten in der Tabelle `co6_data_string`. Zu diesem Zweck werden auch PostgreSQL Funktionen für die Validierung des Wertformats und Konversion des Datentyps des Werts in der Spalte `val` der Tabelle `co6_data_string` angewandt.

4.10.1 Modellierung der Zuweisung der Felder beider Systeme

Die Zuweisung der Felder beider Systeme für die Überführung von Biosignaldaten der Type „Decimal“ und „String“ ist wie folgt:

Input:

- Datensätze aus COPRA
 - `co6_config_variables`
 - `co6_medic_patient`
 - `co6_data_decimal_6_3`
 - `co6_data_string`
- Datensatz aus den FHIR-Profile
 - `mapping_mii_co6_2`
- Datensatz aus `information_schema` von PostgreSQL
 - `tables`

Output:

- FHIR-Ressource der Kategorie „Observation“ - numerische und String Werte

Data Mapping (inhaltlich) - numerische und String Werte		
id		MD5-Hash aus der Kombination des Namens der Werttabelle mit den Hauptschlüssel der Werte in dieser Tabelle und des Namens des FHIR-Profils: <code>md5(tables.table_name, co6_data_decimal_6_3.id</code> und <code>mapping_mii_co6_2.profile.name)</code> oder <code>md5(tables.table_name, co6_data_string.id</code> und <code>mapping_mii_co6_2.profile.name)</code>
meta	profile	<code>mapping_mii_co6_2.meta_profile</code>
status		final
category	coding	system: <code>mapping_mii_co6_2.categoriycoding_system</code>
		code: <code>mapping_mii_co6_2.categoriycoding_code</code>
code	coding	system (SNOMED CT-URL): <code>mapping_mii_co6_2.code_coding_system_snomed</code>
		code (SNOMED CT): <code>mapping_mii_co6_2.code_coding_code_snomed</code>
	coding	system (LOINC-URL): <code>mapping_mii_co6_2.code_coding_system_loinc</code>
		code (LOINC): <code>mapping_mii_co6_2.code_coding_code_loinc</code>
	coding	system (IEEE-URN): <code>mapping_mii_co6_2.code_coding_system_ieee</code>
		code (IEEE): <code>mapping_mii_co6_2.code_coding_code_ieee</code>
subject	reference	Pseudonymisierte Patientennummer: <code>co6_medic_patient.patid</code>

valueQuantity	value	Wert der Messung multipliziert mit dem Umrechnungsfaktor: <code>co6_data_decimal_6_3.val</code> mal <code>mapping_mii_co6_2.unit_transform</code> oder <code>to_numeric(co6_data_string.val)</code> mal <code>mapping_mii_co6_2.unit_transform</code> . Wobei <code>to_numeric()</code> die Funktionen für das type casting darstellen.
	system	<code>mapping_mii_co6_2.valueQuantity_system</code>
	code	Maßeinheit: <code>mapping_mii_co6_to_transfer.profile_unit</code>
effectiveDateTime		Datum und Uhrzeit der Messung: <code>co6_data_decimal_6_3.datetimeto</code> oder <code>co6_data_string.datetimeto</code>

Die Zuweisung der Felder beider Systeme für die Überführung der Biosignalparameter von Blutdruckmessungen ist wie folgt:

Input:

- Datensätze aus COPRA
 - `co6_config_variables`
 - `co6_medic_patient`
 - `co6_medic_pressure`
- Datensatz aus den FHIR-Profilen
 - `mapping_mii_co6_2`
- Datensatz aus `information_schema` von PostgreSQL
 - `tables`

Output:

- FHIR-Ressource der Kategorie „Observation“ - Blutdruckmessungen

Data Mapping (inhaltlich) - Blutdruckmessungen		
id		MD5-Hash aus der Kombination des Namens der Werttabelle mit den Hauptschlüssel der Werte in dieser Tabelle und des Namens des FHIR-Profiles: <code>md5(tables.table.name, co6_data_medic_pressure.id und mapping_mii_co6_2.profile.name)</code>
meta	profile	<code>mapping_mii_co6_2.meta.profile</code>
status		final
category	coding	<code>system: mapping_mii_co6_2.category_coding_system</code>
		<code>code: mapping_mii_co6_2.category_coding_code</code>
subject	reference	Pseudonymisierte Patientenummer: <code>co6_medic_patient.patid</code>
effectiveDateTime		Datum und Uhrzeit der Messung: <code>co6_medic_pressure.datetimet</code>
component		
code (Systolisch)	coding	<code>system (SNOMED CT-URL): mapping_mii_co6_2.code_systolic_coding_system_snomed</code>
		<code>code (SNOMED CT): mapping_mii_co6_2.code_systolic_coding_code_snomed</code>
	coding	<code>system (LOINC-URL): mapping_mii_co6_2.code_systolic_coding_system_loinc</code>
		<code>code (LOINC): mapping_mii_co6_2.code_systolic_coding_code_loinc</code>
	coding	<code>system (IEEE-URN): mapping_mii_co6_2.code_systolic_coding_system_ieee</code>
		<code>code (IEEE): mapping_mii_co6_2.code_systolic_coding_code_ieee</code>
valueQuantity	value	Systolischer Wert: <code>co6_medic_pressure.systolic</code>
	system	<code>http://unitsofmeasure.org</code>

	code	Maßeinheit: mapping_mii_co6.to_transfer.profile_unit
code (Mittel)	coding	system (SNOMED CT-URL): mapping_mii_co6_2.code_mean_coding_system_snomed
		code (SNOMED CT): mapping_mii_co6_2.code_mean_coding_code_snomed
	coding	system (LOINC-URL): mapping_mii_co6_2.code_mean_coding_system_loinc
		code (LOINC): mapping_mii_co6_2.code_mean_coding_code_loinc
	coding	system (IEEE-URN): mapping_mii_co6_2.code_mean_coding_system_ieee
		code (IEEE): mapping_mii_co6_2.code_mean_coding_code_ieee
valueQuantity	value	Systolischer Wert: co6_medic_pressure.mean
	system	http://unitsofmeasure.org
	code	Maßeinheit: mapping_mii_co6.to_transfer.profile_unit
code (Diastolisch)	coding	system (SNOMED CT-URL): mapping_mii_co6_2.code_diastolic_coding_system_snomed
		code (SNOMED CT): mapping_mii_co6_2.code_diastolic_coding_code_snomed
	coding	system (LOINC-URL): mapping_mii_co6_2.code_diastolic_coding_system_loinc
		code (LOINC): mapping_mii_co6_2.code_diastolic_coding_code_loinc
	coding	system (IEEE-URN): mapping_mii_co6_2.code_diastolic_coding_system_ieee
		code (IEEE): mapping_mii_co6_2.code_diastolic_coding_code_ieee
valueQuantity	value	Diastolischer Wert: co6_medic_pressure.diastolic

	system	http://unitsofmeasure.org
	code	Maßeinheit: mapping_mii_co6_to_transfer.profile_unit

Mit den vorherigen Spezifikationen für die Überführung der Daten wurden danach die SQL-Views in der COPRA-Instanz des Staging Bereichs des DWH programmiert, um die Daten der COPRA-Instanz für die Überführung in die FHIR-Ressourcen bereitzustellen.

4.11 Bereitstellung der Biosignaldaten

Mit der Modellierung der Zuweisung der Felder und der Definition von neuen Transformationsregeln für die Überführung der Biosignaldaten aus COPRA in die FHIR-Ressourcen des Moduls „Intensivmedizin“ wurden drei SQL-Views programmiert, um die Elemente der Profile, die Attribute der Biosignaldaten in den Werttabellen, zusammen mit den dazugehörigen Daten der behandelnden Personen, zu verlinken und zu visualisieren. Die angelegten SQL-Views sind folgende:

- `v_profil_decimal`: Information der Profile und der Biosignaldaten in der Tabelle `co6_data_decimal_6_3`
- `v_profil_string`: Information der Profile und der Biosignaldaten in der Tabelle `co6_data_string`
- `v_profil_pressure`: Information der Profile und der Biosignaldaten in der Tabelle `co6_medic_pressure`

Der Inhalt der Spalten der SQL-Views sind die spezifizierten Parameter der Tabellen in der Subsektion 4.10.1.

Der Code 4.6 zeigt den SQL-Befehl für die Erzeugung der SQL-View für die Bioparameter in der Tabelle `co6_data_string`.

```
create or replace view copra.v_profil_string
as
select
    's_' || md5(
        (select table_name
         from information_schema.tables
         where table_schema = 'copra'
         and table_name = 'co6_data_string')
        || cdd.id
```

```

    || mmc.profile_name) id, -- ID aus der Zusammensetzung vom Name
der Werttabelle, id des Werts und name des Profils
    mmc.meta_profile,
    'final' status,
    mmc.category_coding_system,
    mmc.category_coding_code,
    mmc.code_coding_system_snomed,
    mmc.code_coding_code_snomed,
    mmc.code_coding_system_loinc,
    mmc.code_coding_code_loinc,
    mmc.code_coding_system_ieee,
    mmc.code_coding_code_ieee,
    'p_' || md5(cmdp.id::varchar) subject_reference,
    cdd.val::decimal * mmc.unit_transform "valueQuantity_value", -- type
casting und Umrechnung
    mmc.valuequantity_system "valueQuantity_system",
    mmc.valuequantity_code "valueQuantity_code",
    cdd.datetimeto "effectiveDateTime"
from copra.co6_data_string cdd
join copra.co6_config_variables ccv
    on cdd.varid = ccv.id
join copra.mapping_mii_co6_2 mmc
    on mmc.conf_var_id = ccv.id
join copra.co6_medical_data_patient cmdp
    on cmdp.id = cdd.parent_id
where not cdd.deleted
and cdd.validated
and cdd.flagcurrent
and cdd.val ~ '^d+$|^d+\.d+$' -- Kontrolle der numerischen Struktur
der Werte
;

```

Listing 4.5: SQL-View für Werte in co6_data_string.

Die SQL-View für die Biosignale in co6_data_decimal_6.3 wird in dieser Arbeit nicht präsentiert, denn diese View ist ähnlich wie die View für die String-Werte, aber weniger komplex. Der Grund hierzu ist, dass die detektierten Biosignaldaten in co6_data_string, die den FHIR-Profilen zugeordnet sind, in Wahrheit numerische Einträge sind (Tabelle 4.26), und somit sollte der Datentyp der Werte der Biosignaldaten umgewandelt werden, jedoch muss zuvor die Struktur des Wertes in der Spalte val der Tabelle co6_data_string kontrolliert werden (4.10).

Tabelle 4.26: Beispiel eines Eintrags in der Tabelle `co6_data_string`. Der angezeigte Wert des Biosignals ist eine Zahl.

Profil	Konfigurationsvariable	Wert
Linksventrikulaeres Schlagvolumen	SV	78.1

In der SQL-View für die Blutdruckwerte (Code 4.6) werden die Suffixe „systolic“ für systolisch, „mean“ für mittel und „diastolic“ für diastolisch angewendet, um die Zusammensetzung dieser Werte in den zugeordneten FHIR-Profilen darzustellen (4.2).

```
create or replace view copra.v_profil_pressure
as
select
  'pr_' || md5(
    (select table_name
     from information_schema.tables
     where table_schema = 'copra'
     and table_name = 'co6_medic_pressure')
    || cdd.id
    || mmc.profile_name) id,
  mmc.meta_profile,
  'final' status,
  mmc.category_coding_system,
  mmc.category_coding_code,
  'p_' || md5(cmdp.id::varchar) subject_reference,
  cdd.datetimeteto "effectiveDateTime",
  mmc.code_systolic_coding_system_snomed,
  mmc.code_systolic_coding_code_snomed,
  mmc.code_systolic_coding_system_loinc,
  mmc.code_systolic_coding_code_loinc,
  mmc.code_systolic_coding_system_ieee,
  mmc.code_systolic_coding_code_ieee,
  cdd.systolic "valueQuantity_value_systolic",
  mmc.valuequantity_system "valueQuantity_system_systolic",
  mmc.valuequantity_code "valueQuantity_code_systolic",
  mmc.code_mean_coding_system_snomed,
  mmc.code_mean_coding_code_snomed,
  mmc.code_mean_coding_system_loinc,
  mmc.code_mean_coding_code_loinc,
  mmc.code_mean_coding_system_ieee,
  mmc.code_mean_coding_code_ieee,
  cdd.mean "valueQuantity_value_mean",
  mmc.valuequantity_system "valueQuantity_system_mean",
  mmc.valuequantity_code "valueQuantity_code_mean",
  mmc.code_diastolic_coding_system_snomed,
  mmc.code_diastolic_coding_code_snomed,
  mmc.code_diastolic_coding_system_loinc,
```

```

mmc.code_diastolic_coding_code_loinc,
mmc.code_diastolic_coding_system_ieee,
mmc.code_diastolic_coding_code_ieee,
cdd.mean "valueQuantity_value_diastolic",
mmc.valuequantity_system "valueQuantity_system_diastolic",
mmc.valuequantity_code "valueQuantity_code_diastolic"
from copra.co6_medic_pressure cdd
join copra.co6_config_variables ccv
  on cdd.varid = ccv.id
join copra.mapping_mii_co6_2 mmc
  on mmc.conf_var_id = ccv.id
join copra.co6_medic_data_patient cmdp
  on cmdp.id = cdd.parent_id
where cdd.validated
and not cdd.deleted
and cdd.flagcurrent
;

```

Listing 4.6: SQL-View für Werte in co6_medic_pressure.

Mit diesen SQL-Views können FHIR-Ressourcen wie im Code 4.7 erzeugt werden, denn eine Zeile in der View entspricht einer FHIR-Ressource im Modul „Intensivmedizin“.

```

{
  "resourceType": "Observation",
  "id": "pr_d43de0fb06ce5b34d7692ad69b59771f",
  "meta": {
    "profile": [
      "https://medizininformatik-initiative.de/fhir/ext/modul-icu/
      StructureDefinition/pulmonalarterieller-blutdruck"
    ]
  },
  "status": "final",
  "category": [
    {
      "coding": [
        {
          "system": "http://terminology.hl7.org/
          CodeSystem/observation-category",
          "code": "vital-signs"
        }
      ]
    }
  ],
  "code": {
    "coding": [
      {
        "system": "http://snomed.info/sct",

```

```

        "code": "75367002",
      }
    ]
  },
  "subject": {
    "reference": "p_e139c454239bfde741e893edb46a06cc"
  },
  "effectiveDateTime": "2022-11-25T09:30:11+01:00",
  "component": [
    {
      "code": {
        "coding": [
          {
            "system": "http://loinc.org",
            "code": "8406-1"
          },
          {
            "system": "urn:iso:std:iso:11073:10101",
            "code": "150107"
          }
        ]
      },
      "valueQuantity": {
        "value": 104,
        "system": "http://unitsofmeasure.org",
        "code": "mm[Hg]"
      }
    },
    {
      "code": {
        "coding": [
          {
            "system": "http://loinc.org",
            "code": "8432-7"
          },
          {
            "system": "urn:iso:std:iso:11073:10101",
            "code": "150105"
          }
        ]
      },
      "valueQuantity": {
        "value": 86,
        "system": "http://unitsofmeasure.org",
        "code": "mm[Hg]"
      }
    }
  ],
  {
    "code": {

```



```

        "coding": [
          {
            "system": "http://loinc.org",
            "code": "8377-4"
          },
          {
            "system": "urn:iso:std:iso:11073:10101",
            "code": "150106"
          }
        ],
        "valueQuantity": {
          "value": 67,
          "system": "http://unitsofmeasure.org",
          "code": "mm[Hg]"
        }
      }
    ]
  }

```

Listing 4.7: Beispiel einer FHIR-Ressource einer Blutdruckmessung aus COPRA.

Ergebnisse und Diskussion

Nach der vorherigen Erläuterung der Realisierung und deren Ereignisse, erfolgt in diesem Kapitel eine kritische Bewertung der Resultate dieses Projekts. Das Kapitel ist in drei Sektionen gegliedert. Zuerst wird das Hauptergebnis dieser Arbeit diskutiert, nämlich die Zuordnung der Biosignaldaten aus COPRA mit den FHIR-Profilen. Die zweite Sektion befasst sich mit den Parametern der FHIR-Profile. In der dritten Sektion werden die Biosignaldaten von COPRA und deren Einfluss auf die Ergebnisse dieses Projekts diskutiert.

5.1 Zuordnung der COPRA-Biosignaldaten mit den FHIR-Profilen des Moduls „Intensivmedizin“

Das Hauptziel dieser Arbeit wurde durch die ersten Schritten des Data Mappings der Biosignaldaten aus COPRA mit den FHIR-Spezifikationen des Erweiterungsmoduls „Intensivmedizin“ erreicht. Innerhalb dieser Schritte ist ein Datensatz mit der Zuordnung der COPRA-Biosignaldaten mit den FHIR-Profilen entstanden. Mit Hilfe von diesem Datensatz wurden am Ende SQL-Views in die COPRA-Instanz des Staging Bereichs des DWH des DIZ programmiert. Solche virtuellen Tabellen dienen der Zusammenführung und Bereitstellung der Parameter beider Systeme und können somit das Gerüst für die FHIR-Ressourcen sein.

Am Ende dieses Projekts wurden 70 Biosignaldaten vom COPRA-System richtig 39 FHIR-Profilen zugeordnet. Dieses Ergebnis wurde in einem Datensatz mit 75 Einträgen registriert.

Die meisten Werte der Bioparameter in COPRA sind der Typ Dezimal (64), wie die Abbildung 5.1 zeigt. Dieser Datentyp stimmt mit dem spezifizierten Datentyp des Attributs `val` des Elements `valueQuantity` in

den FHIR-Profilen überein. Auch der Aufbau der Biosignaldaten für die Blutdruckmessungen in COPRA stimmt mit der Struktur der FHIR-Profile überein. Nur eine Konfigurationsvariable war der Typ String (Tabelle 4.26) aber mit registrierten numerischen Werten, die bei der Programmierung der Transformationsregeln umgewandelt wurde (4.11).

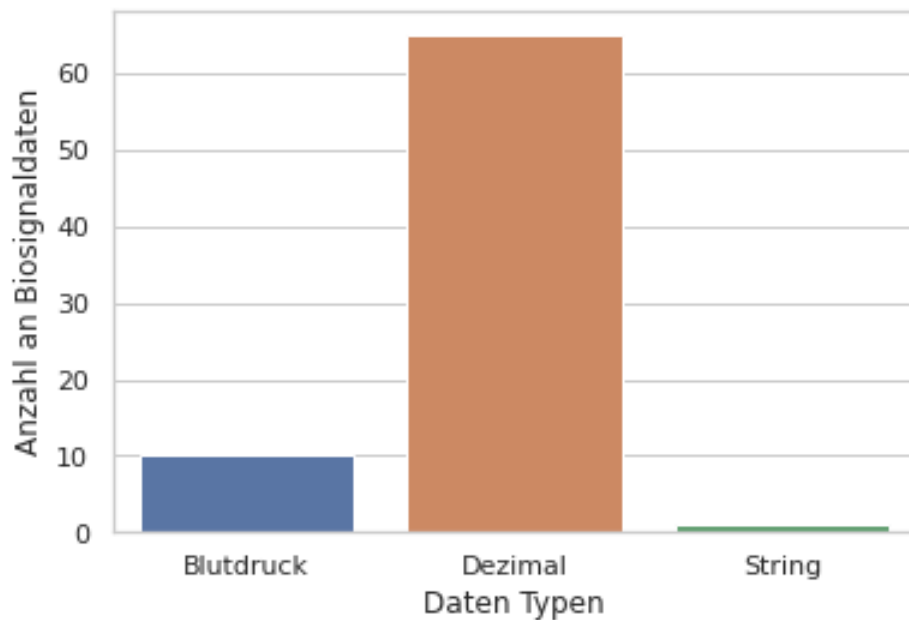


Abbildung 5.1: Datentypen der Biosignaldaten.

Auch die Distribution der Datenarten der FHIR-Profile ist zu berücksichtigen (Abbildung 5.2). Die meist repräsentierte Datenart der Profile im Datensatz der Zuordnung ist die Art „Parameter von Beatmung“ mit 36 zugeordneten Biosignaldaten. Diese Datenart ist auch die meist repräsentierte Art im Erweiterungsmodul „Intensivmedizin“.



Abbildung 5.2: Datenarten der Profile.

5.2 FHIR-Profile

Obwohl die FHIR-Profile des Moduls „ICU“ für die Interoperabilität dienen sollten, beinhalten sie während der Durchführung dieses Projekts einige Irregularitäten, da das Modul in dieser Zeit in Abstimmung sich befindet, und weitere Untersuchungen und Tests von den MII-Mitgliedern und anderen Benutzern erfolgen. Aus diesem Grund beinhalten zwei FHIR-Profile im Erweiterungsmodul denselben Namen, aber spezifizieren unterschiedlichen Beobachtungen.

Die meisten definierten FHIR-Profile gehören zu der Kategorie „Observation“ (Tabelle 4.2). Das liegt daran, dass das Ziel des Moduls die Spezifikationen der akutmedizinischen Daten ist, und gerade im Bereich der Intensivmedizin verlaufen die meisten Beobachtungen und Messungen („Observation“).

Ein wichtiger Parameter für die Einhaltung der Interoperabilität im Gesundheitswesen sind die Codesysteme (Tabelle 4.3). Alle nicht generischen Profile beinhalten zumindest ein standardisiertes Codesystem. Die Profile der Kategorie „Observation“ ohne Codesysteme gehören zu den generischen Profilen (Abbildung 5.3) und dienen der Modellierung der Daten (2.2.1). Aus diesem Grund beinhalten solche Profile keine Codesysteme und auch keine

Maßeinheiten (Tabelle 4.4). Die Profile der Kategorie „Procedure“ beinhalten auch keine Codesysteme, denn diese Profile speichern die Information der angewandten Geräte für die Durchführung eines Verfahrens und diese Information befindet sich im Basismodul „Prozedur“ der Kerndatensatz der MII.



Abbildung 5.3: Baumstruktur des Erweiterungsmoduls „Intensivmedizin“. Die generischen Profile der Kategorie „Observation“ sind rot markiert.

Das Modul „ICU“ mit 80 FHIR-Profilen ist das umfangreichste Modul der MII. Von diesen Profilen könnte in diesem Projekt 39 einem Biosignalparameter zugeordnet werden (Abbildung 5.4). Die Gründe dieses Ergebnisses sind nicht nur von dem Umfang des Moduls abhängig, sondern von den Bio-parameter im COPRA-System (5.3).



Abbildung 5.4: Diagramm der FHIR-Profile im Projekt. Fast die Hälfte der FHIR-Profile konnten zugeordnet werden.

5.3 Biosignaldaten aus COPRA

Die Daten aus COPRA zeigten eine sehr geringe Interoperabilität. Ein Grund hierzu ist die fehlende Standardisierung bei der Schreibweise und Art der Maßeinheiten, da manche Einheiten und deren Darstellung nicht durch internationale etablierte Standards, wie die UCUM definiert sind. Das hat zur Folge, dass die Maßeinheiten zwischen beiden Systemen harmonisiert werden müssen. Das hat zur Folge, dass ein Teil der Information von COPRA verloren geht. Für die Überführung der Daten aus COPRA sollen die spezifizierten Einheiten der FHIR-Profile benutzt werden, denn die Maßeinheiten in den FHIR-Profilen sind durch standardisierte Codesysteme definiert.

Darüber hinaus sind manche Biosignaldaten im COPRA-System nicht harmonisiert, denn das COPRA-System beinhaltet keine der etablierten Codesysteme für die Codierung der Bioparameter. Das hat zur Folge, dass in einigen Situationen dasselbe Biosignal von unterschiedlichen Geräten und/oder in diversen Stationen gemessen wird, aber unterschiedlich im System wahrgenommen wird. Ein Beispiel davon befindet sich in der Tabelle 5.1.

Tabelle 5.1: Dasselbe Biosignal gemessen durch zwei Geräte.

Profilname	COPRA-Name	LOINC	SNOMED CT	Einheit
Beatmungsvolumen-Pro-Minute-Machineller-Beatmung	Beatmung _MS_Evita4 _MV	76009-0	426102006	L/min
Beatmungsvolumen-Pro-Minute-Machineller-Beatmung	Beatmung _MS_Pallas _MV	76009-0	426102006	L/min

Nur eine der zugeordneten Biosignaldaten ist der Typ String (Tabelle 4.26). Durch andere Projekte, wie der Aktionsbündnis für Informations- und Kommunikationstechnologie in Intensiv- und Notfallmedizin (AKTIN)-Notaufnahmeregister, wurde zur Kenntnis genommen, dass es weitere Biosignaldaten, wie die Pupillenweite oder Pupillenreaktion, auch der Datentyp String sind, und deren Informationen in den aktuellen COPRA-Tabellen des Staging Bereichs des DWH sich befinden. Diese anderen Biosignalparameter können bei der Weiterentwicklung des Moduls „Intensivmedizin“ betrachtet werden.

Andere erkannte Unregelmäßigkeiten sind die Namen und die Beschreibungen der Konfigurationsvariablen in COPRA. Einerseits beinhalten viele Biosignaldaten keine Beschreibung für die Fachtermini im System, die die Daten dokumentieren. In dieser Situation befinden sich 379 der 701 relevanten Konfigurationsvariablen. Andererseits verfolgt die Benennung der Konfigurationsvariablen keine Standardnomenklatur, weder auf Deutsch noch auf Englisch. Aus diesem Grund wurden viele FHIR-Profile nicht gefunden. Aus diesem Grund ist viel Information verloren gegangen.

Einige Gründe für diese Problematiken sind zu einem die unterschiedlichen genutzten Geräte und zum anderen historische Hintergründe, da das COPRA-System erst seit 2007 an der Universitätsmedizin Mainz etabliert und von verschiedenen Anwender und Anwenderinnen bedient wurde. Hinzukommend werden viele Daten im COPRA-System manuell und nicht automatisiert eingetragen. Es gibt beispielsweise numerische Werte, die in Feldern für Freitexte gespeichert werden. Somit leidet die Qualität und Brauchbarkeit der Daten.

Einige FHIR-Profile wurden nicht gefunden, denn manche Beobachtungen oder Messungen am Standort Mainz werden im COPRA nicht erfasst. Ein Beispiel davon ist das Profil „Horowitz-In-Arteriell-Blut“ in mm[Hg].

Stattdessen wird in COPRA der Horowitz-Index ohne Einheit registriert.

Trotz dieser Hürden, konnten 70 Konfigurationsvariablen einem FHIR-Profil zugeordnet werden (Abbildung 5.5). Die Ursachen dieses Ergebnisses liegen an den vorher genannten Problematiken und an der Komplexität und dem Umfang des Erweiterungsmoduls „Intensivmedizin“.

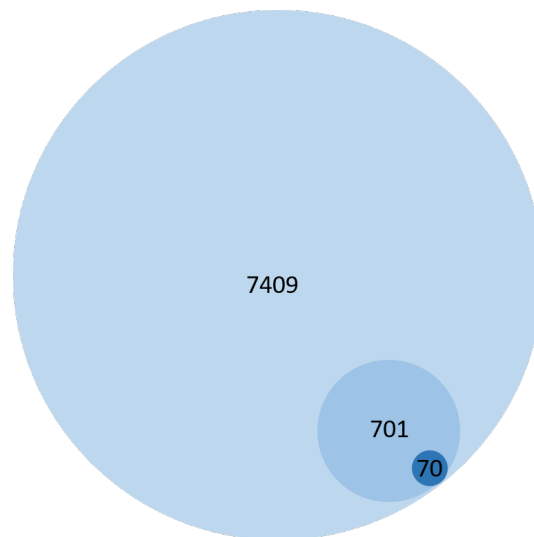


Abbildung 5.5: Diagramm der Konfigurationsvariablen im Projekt. 70 der 701 relevanten Konfigurationsvariablen entsprechen einem Biosignalparameter, welches einem FHIR-Profil zugeordnet wurde.

Fazit und Ausblick

Mit der Zuordnung der gespeicherten Biosignaldaten aus dem PDMS mit den FHIR-Spezifikationen des Erweiterungsmoduls „Intensivmedizin“ und die Bereitstellung dieser Daten für die Überführung in FHIR wurde das Ziel dieser Arbeit erreicht. Somit entstand ein Grundgerüst, welches für die Überführung der Biosignaldaten in FHIR-Ressource benutzt werden konnte. Diese Schritte sind entscheidend für die Gewährleistung der Interoperabilität der Biosignaldaten in COPRA, und sind somit notwendige Komponenten der Datenmigration und Datenintegration.

Die Relevanz dieses Projekts liegt daran, dass die Biosignalparameter aus dem COPRA-System der Universitätsmedizin Mainz mit Hilfe von der praktischen Umsetzung des Data Mappings in ein Standardformat überführt werden können. Das hat zur Folge, dass diese Daten für Forschungszwecke weiterbenutzt werden können. Damit können die standardisierten Daten unter anderem für Datenauswertungen und Reports für Anwendungszwecke, Statistik, oder die Entwicklung von Systemen für die Unterstützung von klinischen Entscheidungen angewendet werden.

Ein weiterer wichtiger Aspekt dieses Projekts ist, dass mit der Überführung der Biosignaldaten in ein Standardformat mit standardisierten Codesystemen, die Daten am Ende auch die Findable Accessible Interoperable Reusable (FAIR)-Prinzipien erhalten werden.

Dieses Projekt könnte die Motivation für den Ideenaustausch und die Kooperation mit anderen MII-Standorten steigern.

Literatur

- [1] Alexander Zautke, Berke Dincel, Christoph Müller, Felix Erdfelder, Lily Wissing, Larissa Neumann, Ludwig C. Hinske. *Medizininformatik-Initiative - Modul ICU*. 2022. <https://simplifier.net/medizininformatikinitiative-modul-intensivmedizin> (besucht am 02.06.2022).
- [2] IEEE. „IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries“. In: *IEEE Std 610.12-1990* (1990).
- [3] Gernot Marx, Rolf Rossaint und Nikolaus Marx. *Telemedizin*. 4. Aufl. Springer, Berlin, Heidelberg, 2021. ISBN: 978-3-662-60611-7. DOI: <https://doi.org/10.1007/978-3-662-60611-7>.
- [4] Stefan Müller und Thomas Lux. *E-Health-Ökonomie*. 1. Aufl. Springer, Wiesbaden, 2017. ISBN: 978-3-658-10788-8. DOI: 10.1007/978-3-658-10788-8.
- [5] Tim Benson und Grahame Grieve. *Principles of Health Interoperability SNOMED CT, HL7 and FHIR*. 3. Aufl. Springer, London, 2016. ISBN: 978-3-319-30370-3. DOI: 10.1007/978-3-319-30370-3.
- [6] Medizininformatik-Initiative. *Über die Initiative*. Bundesministerium für Bildung und Forschung. 2022. <https://www.medizininformatik-initiative.de/de/ueber-die-initiative> (besucht am 06.06.2022).
- [7] Medizininformatik-Initiative. *Der Kerndatensatz der Medizininformatik-Initiative*. Bundesministerium für Bildung und Forschung. 2022. <https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative> (besucht am 15.06.2022).
- [8] Medizininformatik-Initiative. *Datenintegrationszentren*. Bundesministerium für Bildung und Forschung. <https://www.medizininformatik-initiative.de/de/konsortien/datenintegrationszentren> (besucht am 11.06.2022).

- [9] Medizininformatik-Initiative. *Kommentierung: MII-Kerndatensatzmodul Intensivmedizin — PDMS-Daten und hochauflösende Biosignaldaten*. Bundesministerium für Bildung und Forschung. 2022. <https://www.medizininformatik-initiative.de/index.php/de/kommentierung-mii-kerndatensatzmodul-intensivmedizin-pdms-daten-und-hochaufloesende-biosignaldaten> (besucht am 15.06.2022).
- [10] Felix Erdfelder. *Erweiterungsmodul Intensivmedizin der Medizininformatik-Initiative*. [Video MIRACUM-Kolloquium]. 2022. <https://confluence.imi.med.fau.de/pages/viewpage.action?pageId=207390115> (besucht am 15.06.2022).
- [11] FHIR HL7. *DeviceMetric*. 2022. <https://simplifier.net/packages/hl7.fhir.r4.core/4.0.1/files/82244/~details> (besucht am 12.08.2022).
- [12] FHIR HL7. *Observation*. 2022. <https://simplifier.net/packages/hl7.fhir.r4.core/4.0.1/files/83132/~details> (besucht am 12.08.2022).
- [13] FHIR HL7. *Procedure*. 2022. <https://simplifier.net/packages/hl7.fhir.r4.core/4.0.1/files/79221/~details> (besucht am 12.08.2022).
- [14] HL7 International. *About HL7*. HL7®. 2022. <http://www.hl7.org/about/index.cfm?ref=nav> (besucht am 05.06.2022).
- [15] HL7 Deutschland. *Warum FHIR?* HL7 Deutschland e. V. 2022. <https://hl7.de/themen/hl7-fhir-mobile-kommunikation-und-mehr/warum-fhir/> (besucht am 05.06.2022).
- [16] HL7 International. *Change Management and Versioning*. HL7®. Mai 2022. <http://hl7.org/fhir/versions.html> (besucht am 05.06.2022).
- [17] HL7 International. *Formats*. HL7®. Mai 2022. <https://www.hl7.org/fhir/formats.html> (besucht am 06.06.2022).
- [18] Hartmut Ernst, Jochen Schmidt und Gerd Beneken. *Grundkurs Informatik*. 7. Aufl. Springer, Wiesbaden, 2020. ISBN: 978-3-658-30331-0. DOI: <https://doi.org/10.1007/978-3-658-30331-0>.
- [19] Quality Services und Wissen GmbH. *XML – Extensible Markup Language*. Quality Services & Wissen GmbH. 2022. <https://www.quality.de/lexikon/xml/> (besucht am 15.06.2022).
- [20] ORACLE. *Was ist JSON?* ORACLE. 2022. <https://www.oracle.com/de/database/what-is-json/> (besucht am 16.06.2022).

- [21] Centers for Medicare & Medicaid Services. *Codes, Code Systems, and Value Sets*. Centers for Medicare & Medicaid Services. 2021. <https://www.cms.gov/files/document/blueprint-codes-code-systems-value-sets.pdf> (besucht am 16.06.2022).
- [22] HL7[®] FHIR[®]. *Resource CodeSystem - Content*. HL7[®] FHIR[®]. 2019. <https://www.hl7.org/fhir/codesystem.html> (besucht am 16.06.2022).
- [23] SNOMED International. *SNOMED International*. SNOMED International. 2022. <https://www.snomed.org/> (besucht am 20.06.2022).
- [24] SOMED International. *SNOMED CT Start Guide*. SOMED International. 2012. <https://confluence.ihtsdotools.org/display/DOCSTART/SNOMED+CT+Starter+Guide> (besucht am 20.06.2022).
- [25] Bundesinstitut für Arzneimittel und Medizinprodukte. *LOINC*. Bundesministerium für Gesundheit. 2022. https://www.bfarm.de/DE/Kodiersysteme/Terminologien/LOINC-UCUM/LOINC-und-RELMA/_node.html (besucht am 20.06.2022).
- [26] Michelle Stram, Tony Gigliotti, Douglas Hartman, Andrea Pitkus, Stanley M. Huff, Michael Riben, Walter H. Henricks, Navid Farahani und Liron Pantanowitz. *Logical Observation Identifiers Names and Codes for Laboratorians: Potential Solutions and Challenges for Interoperability*. Juni 2019. DOI: 10.5858/arpa.2018-0477-RA. eprint: https://meridian.allenpress.com/aplm/article-pdf/144/2/229/2500745/arpa_2018-0477-ra.pdf. <https://doi.org/10.5858/arpa.2018-0477-RA>.
- [27] LOINC. *Basics*. LOINC. 2021. <https://loinc.org/kb/faq/basics/> (besucht am 25.10.2022).
- [28] LOINC. *Structure*. LOINC. 2021. <https://loinc.org/kb/faq/structure/> (besucht am 25.10.2022).
- [29] IEE. *ISO/IEEE 11073-10101:2020(en) Health informatics — Device interoperability — Part 10101: Point-of-care medical device communication — Nomenclature*. IEEE. 2019. <https://www.iso.org/obp/ui#iso:std:iso-ieee:11073:-10101:ed-2:v1:en> (besucht am 21.06.2022).
- [30] IEEE. *IEEE at a Glance*. IEEE. 2022. <https://www.ieee.org/about/at-a-glance.html> (besucht am 21.06.2022).
- [31] Stefan Schlichting und Stephan Pöhlens. *An Architecture for distributed systems of medical devices in high acuity environments*. Jan. 2014.

- [32] IEE. *IEEE/ISO/IEC International Standard - Health informatics-Device interoperability-Part 10101: Point-of-care medical device communication-Nomenclature*. IEEE. Aug. 2020. <https://standards.ieee.org/ieee/11073-10101/10343/> (besucht am 21.06.2022).
- [33] Björn Andersen, Martin Kasparick, Frank Golasowski und Josef Ingnerf. „Extending the IEEE 11073-1010X nomenclature for the modelling of surgical devices“. In: *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE. 2016, S. 244–247.
- [34] U.S. National Library of Medicine. *Unified Code for Units of Measure (UCUM)*. 2022. <https://ucum.nlm.nih.gov/> (besucht am 20.10.2022).
- [35] Unified Code for Units of Measure. *What is UCUM?* 2022. <https://ucum.org/about> (besucht am 20.10.2022).
- [36] Ixchel Castellanos, Jürgen Schüttler, Hans-Ulrich Prokosch und Thomas Bürkle. „Does introduction of a Patient Data Management System (PDMS) improve the financial situation of an intensive care unit?“ In: *BMC Medical Informatics and Decision Making* 13.1 (Sep. 2013), S. 107. ISSN: 1472-6947. DOI: 10.1186/1472-6947-13-107. <https://doi.org/10.1186/1472-6947-13-107> (besucht am 24.06.2022).
- [37] I Castellanos, T Ganslandt, H U Prokosch, J Schüttler und T Bürkle. „Einführung eines Patientendatenmanagementsystems“. In: *Der Anaesthesist* 62.11 (Nov. 2013), S. 887–90, 892–7. ISSN: 1432-055X. DOI: 10.1007/s00101-013-2239-x. <https://doi.org/10.1007/s00101-013-2239-x> (besucht am 24.06.2022).
- [38] G. Daniel Martich, Carl S. Waldmann und Michael Imhoff. „Clinical Informatics in Critical Care“. In: *Journal of Intensive Care Medicine* 19.3 (2004). PMID: 15154996, S. 154–163. DOI: 10.1177/0885066604264016. eprint: <https://doi.org/10.1177/0885066604264016>. <https://doi.org/10.1177/0885066604264016>.
- [39] Heiko Mania. *PDMS : Marktübersicht*. Health.Tech.Blog. Mai 2021. <https://gesundheitstechnologie.online/wiki/pdms-marktuebersicht/> (besucht am 27.06.2022).
- [40] Dosing GmbH. *COPRA SYSTEM GMBH*. Dosing GmbH. 2022. <https://www.dosing-gmbh.de/partner/copra-system-gmbh/> (besucht am 24.06.2022).
- [41] COPRA System GmbH. *Copra*. COPRA System GmbH. 2022. <https://copra-system.de/produkt/> (besucht am 27.06.2022).

- [42] COPRA. *Referenzen*. COPRA. 2022. <https://copra-system.de/referenzen/> (besucht am 27.10.2022).
- [43] COPRA System GmbH. *COPRA Datenbank Beschreibung*. Techn. Ber. COPRA System GmbH, Okt. 2015.
- [44] William H Inmon. *Building the Data Warehouse*. John Wiley & Sons, 2005. ISBN: 978-0-7645-9944-6.
- [45] Shawn Murphy. „Data Warehousing for Clinical Research“. In: *Encyclopedia of Database Systems*. Hrsg. von LING LIU und M. TAMER ÖZSU. Boston, MA: Springer US, 2009, S. 679–684. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_120. https://doi.org/10.1007/978-0-387-39940-9_120.
- [46] Matthias Goeken. *Entwicklung von Data-Warehouse-Systemen: Anforderungsmanagement, Modellierung, Implementierung*. 1. Aufl. Deutscher Universitäts-Verlag, 2006. ISBN: 978-3-8350-0325-5.
- [47] Frank Navrade. *Strategische Planung mit Data-Warehouse-Systemen*. Springer-Verlag, 2008. ISBN: 978-3-8349-9762-3. DOI: <https://doi.org/10.1007/978-3-8349-9762-3>.
- [48] Andreas Bauer und Holger Günzel. *Data-warehouse-Systeme: Architektur, Entwicklung, Anwendung*. dpunkt-Verlag, 2004. ISBN: 9783898642514.
- [49] Ralph Kimball und Margy Ross. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. Aufl. Wiley, 2013. ISBN: 978-1-118-73219-9.
- [50] Qamar Shahbaz. *Data Mapping for Data Warehouse Design*. Computer Science Reviews and Trends. Elsevier Science, 2015. ISBN: 9780128053355. <https://books.google.de/books?id=pRChCgAAQBAJ>.
- [51] Nida Fatima. *Verstehen von Tools, Prozessen und Techniken zur Datenzuordnung*. Astera Software. Dez. 2018. <https://www.astera.com/de/type/blog/understanding-data-mapping-and-its-techniques/> (besucht am 05.07.2022).
- [52] Padmaja Potineni. *Data Warehousing Guide*. ORACLE. 2021. <https://docs.oracle.com/en/database/oracle/oracle-database/21/dwhsg/index.html> (besucht am 30.06.2022).
- [53] David Tylor. *Data Modelling: Conceptual, Logical, Physical Data Model Types*. Guru99. Feb. 2022. <https://www.guru99.com/data-modelling-conceptual-logical.html> (besucht am 30.06.2022).

- [54] New Media Service. *SME – Subject Matter Expert*. New Media Service. 2022. <https://nms.de/it-lexikon/artikel/sme-subject-matter-expert.html> (besucht am 05.07.2022).
- [55] Talend. *Was ist Daten-Mapping?* Talend. 2022. <https://www.talend.com/de/resources/data-mapping/> (besucht am 05.07.2022).
- [56] Alberto Apostolico und Zvil Galil. *Pattern Matching Algorithms*. Oxford University Press, 1997. ISBN: 9780195354348. https://books.google.de/books?id=mFd%5C_grFyiT4C (besucht am 10.07.2022).
- [57] Elmar Cohors-Fresenborg. „Worterzeugende Kalküle“. In: *Mathematik mit Kalkülen und Maschinen*. Wiesbaden: Vieweg+Teubner Verlag, 1977, S. 67–70. ISBN: 978-3-322-85927-3. DOI: 10.1007/978-3-322-85927-3_18. https://doi.org/10.1007/978-3-322-85927-3_18.
- [58] Alfred V. Aho und Margaret J. Corasick. „Efficient String Matching: An Aid to Bibliographic Search“. In: *Commun. ACM* 18.6 (Juni 1975), S. 333–340. ISSN: 0001-0782. DOI: 10.1145/360825.360855. <https://doi.org/10.1145/360825.360855>.
- [59] Nimisha Singla und Deepak Garg. „String matching algorithms and their applicability in various applications“. In: *International journal of soft computing and engineering* 1.6 (2012), S. 218–222.
- [60] Yunyao Li, Rajasekar Krishnamurthy, Sriram Raghavan, Shivakumar Vaithyanathan und H.V. Jagadish. „Regular Expression Learning for Information Extraction“. In: *Empirical Methods in Natural Language Processing (EMNLP)*. Honolulu, HI, 2008.
- [61] PostgreSQL. *Pattern Matching*. PostgreSQL. 2022. <https://www.postgresql.org/docs/current/functions-matching.html> (besucht am 12.07.2022).
- [62] SeatGeek. *thefuzz*. Mai 2022. <https://github.com/seatgeek/thefuzz> (besucht am 31.10.2022).
- [63] William W Cohen, Pradeep Ravikumar, Stephen E Fienberg u. a. „A Comparison of String Distance Metrics for Name-Matching Tasks.“ In: *IIWeb*. Bd. 3. 2003, S. 73–78.
- [64] Pat Brans. *fuzzy search*. Aug. 2022. <https://www.techtarget.com/whatis/definition/fuzzy-search> (besucht am 31.10.2022).
- [65] Ruslan Mitkov. *The Oxford Handbook of Computational Linguistics*. Oxford Handbooks Series. OUP Oxford, 2004. ISBN: 9780199276349. <https://books.google.de/books?id=yl6AnaKtVakC> (besucht am 12.07.2022).

- [66] Kai Priestersbach. *Regex: Einführung in reguläre Ausdrücke, Syntax und deren Verwendung*. Webmasterpro. Dez. 2021. <https://www.webmasterpro.de/coding/einfuehrung-in-regular-expressions/> (besucht am 12.07.2022).
- [67] LOINC. *71835-3 Oxygen/Gas total [Pure volume fraction] Inhaled gas*. LOINC. 2022. <https://loinc.org/71835-3/> (besucht am 01.09.2022).
- [68] Thomas Studer. *Relationale Datenbanken: Von den theoretischen Grundlagen zu Anwendungen mit PostgreSQL*. 1. Aufl. Springer Vieweg Berlin, Heidelberg, 2016. ISBN: 978-3-662-46571-4. DOI: <https://doi.org/10.1007/978-3-662-46571-4>.
- [69] Edwin Schicker. *Datenbanken und SQL: Eine praxisorientierte Einführung mit Anwendungen in Oracle, SQL Server und MySQL*. 5. Aufl. Springer Vieweg Wiesbaden, 2017. ISBN: 978-3-658-16129-3. DOI: <https://doi.org/10.1007/978-3-658-16129-3>.
- [70] Ashlin Jenifa. *Understanding Difference Between URL, URI and URN*. Geekflare. März 2022. <https://geekflare.com/difference-between-url-uri-and-urn/> (besucht am 04.09.2022).
- [71] Andrea Chiarelli. *URL, URI, URN: What's the Difference?* Auth0 Inc. Apr. 2022. <https://auth0.com/blog/url-uri-urn-differences/> (besucht am 04.09.2022).
- [72] Web Hypertext Application Technology Working Group (WHATWG). *URL, Living Standard*. Web Hypertext Application Technology Working Group (WHATWG). Aug. 2022. <https://url.spec.whatwg.org/> (besucht am 04.09.2022).
- [73] LOINC. *75919-1 Left ventricular Cardiac index*. LOINC. 2022. <https://loinc.org/75919-1/> (besucht am 01.09.2022).
- [74] Daw Khin Po. „Similarity Based Information Retrieval Using Levenshtein Distance Algorithm“. In: *Int. J. Adv. Sci. Res. Eng* 6.04 (2020), S. 06–10.

SQL

Structured Query Language (SQL) ist die wichtigste nicht prozedurale Programmiersprache, um mit einem RDMS zu interagieren [68]. Mit SQL werden die Struktur der Daten in einer DB definiert, modifiziert und die Sicherheitsbedingungen aufgebaut [69]. SQL besitzt eine Data Description Language (DDL) für die Beschreibung und Definition der DB und eine Data Manipulation Language (DML), um die Daten in der DB zu manipulieren [68, 69].

- DDL-Befehle
 - CREATE TABLE: Erzeugung einer Tabelle
 - DROP TABLE: Zum Löschen einer Tabelle
 - CREATE VIEW: Erzeugung einer View oder Sicht (virtuelle Tabelle)
 - GRANT: Zum Gewähren von Zugriffsrechten
- DML-Befehle
 - SELECT: Abfragen
 - UPDATE: Änderung der Daten
 - DELETE: Löschung der Daten
 - INSERT: Zum Einfügen von Daten

URI, URL und URN

Ein Uniform Resource Identifier (URI) ist eine Standardmethode für die Identifikation von Ressourcen in Internet, dieser Standard identifiziert die Ressourcen bei seiner Lokalisation, Name oder beides zusammen [70]. Das Schema eines URI ist: **scheme://authority:/path?query# fragment**, wobei **scheme** das Protokoll darstellt, z. B. „HTTP“ [70, 71]. **//authority** identifiziert die Domainadresse der Ressourcen und ist sogleich aus Benutzername, Host und Proxy aufgebaut [70, 72]. **path** zeigt die komplette Adresse der Ressourcen zusammen mit einer Abfrage-Aktion (**?query**) [71]. **fragment** ist ein Teil der referenzierten Ressourcen [70].

Ein Uniform Resource Locator (URL) ist die benutzte Zeichenkette, um die Ressourcen mit Hilfe von der Lokalisation zu erreichen, und ist somit ein Teil des URI [71]. Die Syntax eines URL ist: **scheme: subdomain/domain-name. Top-level-domain/sub-folder**. Bei einem URL, genau wie bei dem URI, liefert **scheme** die Details des Protokolls [70]. **subdomain** ist ein Teil der Domain, wie z. B. eine Facheinrichtung eines Krankenhauses. **domain-name** ist der Name der Domain, im vorherigen Beispiel der Name des Krankenhauses. **Top-level-domain** ist die Adresse der Domain im Internet, z. B. „.de“ [70]. Das optionale Teil **sub-folder** definiert das Verzeichnis der Ressourcen [70, 71].

Das Uniform Resource Name (URN), genau wie das URL, gehört zu dem URI [70]. Das URN hat die folgende Syntax **urn:nid:nss** [70, 71]. **urn** ist die Spezifikation des Schemas. **nid** ist der „namespace identifier“ und definiert den Namensraum der Ressourcen [70]. Am Ende **nss** „namespace-specific string“ ist der Namensraum der Ressourcen selbst [70]. Dieses Attribut ist optional [71]. In einem URN können mehrere Folgen **nid:nss** vorkommen, wie

urn:iso:std:iso:11073:10101 (2.4.3).

Flussdiagramm

Ein Flussdiagramm ist eines der am häufigsten verwendeten Diagramme. Es beschreibt und stellt einen Prozess, ein System oder einen Algorithmus dar. Flussdiagramme werden von technischen und nichttechnischen Personen in zahlreichen Bereichen verwendet, um komplexe Prozesse zu dokumentieren oder optimieren.

Flussdiagramme werden mit zahlreichen Formen (Abbildung C.1) erstellt. Diese werden durch Verbindungspfeile vernetzt, um den Prozessfluss oder den Ablauf zu definieren.

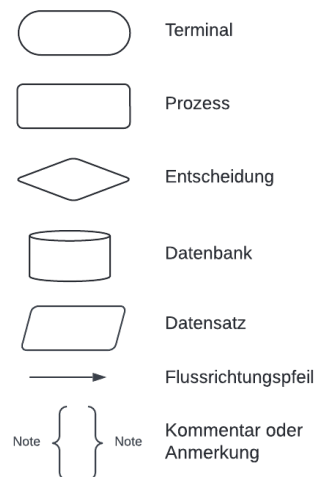


Abbildung C.1: Formen eines Flussdiagramms.

Abbildungsverzeichnis

2.1	Kerndatensatz der MII	8
2.2	Baumstruktur des Erweiterungsmoduls „Intensivmedizin“ . . .	10
2.3	Logische Tabelle eines FHIR-Profiles	13
2.4	Beschreibung der meist benutzten FHIR-Icons	14
2.5	SNOMED CT-Datenmodell	19
2.6	Datenmodell von COPRA	24
2.7	Sternschema	27
3.1	Flussdiagramm des Ablaufes des Projekts	34
4.1	DIZ-Struktur an der Universitätsmedizin Mainz	38
4.2	Komponenten und Fluss der Daten	39
4.3	Zuordnung der Konfigurationsvariablen mit den FHIR-Profilen	58
5.1	Datentypen der Biosignaldaten	78
5.2	Datenarten der Profile	79
5.3	Generische „Observation“-Profile	80
5.4	Diagramm der FHIR-Profile im Projekt	81
5.5	Diagramm der Konfigurationsvariablen im Projekt	83
C.1	Formen eines Flussdiagramms	94

Tabellenverzeichnis

2.1	Blutdruck in SNOMED CT	19
2.2	LOINC-Dimensionen	20
4.1	Struktur der Tabelle mii_icu	41
4.2	FHIR-Profil im Modul „ICU“	44
4.3	Codesysteme der FHIR-Profil im Modul „ICU“	44
4.4	„Observation“-Profil ohne Codesystem	44
4.5	An- und Abwesenheit von Umlauten in den Profilnamen	45
4.6	Relevante Spalten von co6_config_variables	46
4.7	Relevante Spalten von co6_config_variable_types	47
4.8	Relevante Spalten von co6_medic_data_patient	47
4.9	Werttabellen in der COPRA-Instanz	48
4.10	Relevante Spalten von co6_data_decimal_6_3 und co6_data_string	48
4.11	Relevante Spalten von co6_medic_pressure	49
4.12	Anzahl der repräsentierten Konfigurationsvariablen in je Wert- tabelle	51
4.13	Beispiel von LOINC-„Short Name“	52
4.14	Tabelle für die Speicherung der Zuordnung der Konfigurati- onsvariablen mit den FHIR-Profilen	54
4.15	Beispiele von falsch zugeordneten FHIR-Profilen	58
4.16	Erkannte Konfigurationsvariablen durch die Anwendung einer Kombination von allgemeinen Teilwörtern	60
4.17	Erkennung von „Short Names“ von Verfahren in der Tabelle co6_config_variables	61
4.18	Konfigurationsvariablen ohne Maßeinheiten	62
4.19	Schreibweisen derselben Maßeinheiten in beiden Systemen	62
4.20	Beispiel der Schreibweise derselbe Maßeinheit in COPRA	63
4.21	Fehlerhafte Darstellung Maßeinheiten in COPRA	63
4.22	Maßeinheiten und Umrechnungsfaktoren für die Harmonisierung	64
4.23	Inkompatible Maßeinheiten	64
4.26	Eintrag in der Tabelle co6_data_string	73

5.1	Dasselbe Biosignal gemessen durch zwei Geräte	82
-----	---	----

Listingverzeichnis

2.1	Beispiel einer FHIR-Ressource in XML	15
2.2	Beispiel einer FHIR-Ressource in JSON	16
2.3	ISO/IEEE 11073-10101™ in FHIR	21
4.1	SQL-Abfrage zur Auswahl der Konfigurationsvariablen	50
4.2	Python-Skript für das Pattern Matching	57
4.3	SQL-Abfrage mit allgemeinen Teilwörtern	59
4.4	SQL-Abfrage mit einem seltenen Teilwort	60
4.5	SQL-View für Werte in co6_data_string	71
4.6	SQL-View für Werte in co6_medic_pressure	73
4.7	Beispiel einer FHIR-Ressource aus COPRA	74

Abkürzungsverzeichnis

AKTIN	Aktionsbündnis für Informations- und Kommunikationstechnologie in Intensiv- und Notfallmedizin
ANSI	American National Standards Institute
BMBF	Bundesministerium für Bildung und Forschung
COPRA	Computer Organized Patient Report Assistant
CPOE	Computerized Physician Order Entry
CSV	Comma-Separated Values
DB	Datenbank
DDL	Data Description Language
DIZ	Datenintegrationszentrum
DML	Data Manipulation Language
DRG	Diagnosis Related Group
DWH	Data Warehouse
ECMO	Extracorporeal Membrane Oxigenation
ePA	elektronische Patienten Akte
ETL	Extraction, Transformation, Load
FAIR	Findable Accessible Interoperable Reusable
FHIR	Fast Healthcare Interoperability Resources
HL7	Health Level Seven
HTTP	Hypertext Transfer Protocol
ICD	International Statistical Classification of Diseases and Related Health Problems
ICU	Intensive Care Unit
IEEE	Institute of Electrical and Electronics Engineers

IHTSDO International Healthcare Terminology Standards Development Organization

ISO International Organization for Standardization

IT Informationstechnik

JSON JavaScript Object Notation

KIS Krankenhausinformationssystem

LDM Logical Data Model

LOINC Logical Observation Identifiers Names and Codes

LTS Long Term Support

MII Medizininformatik-Initiative

OSI Open Systems Interconnection

PDM Physical Data Model

PDMS Patientendatenmanagementsystem

RDMS Relational Database Management System

RefSets Referenz Sets

Regex Regular Expression

SNOMED CT Systematized Nomenclature of Medicine Clinical Terms

SQL Structured Query Language

XML Extensible Markup Language

XSD Schema Definition Language

UCUM Unified Code for Units of Measure

URI Uniform Resource Identifier

URL Uniform Resource Locator

URN Uniform Resource Name

W3C World Wide Web Consortium

Danksagung

Quiero agradecer muy especialmente a las cuatro mujeres más importantes en mi vida, Fefita, mi mamá, Sandrita y Mici por su amor, confianza y apoyo en todo momento.

Vielen Dank meinem Betreuer, Daniel Schmitz, mein bestes F1 von Anfang bis zum Ende dieses Projekts. Danke an Dr. Paul Schmücker für die Anmerkungen. Vielen Dank an Eric Schultheis, Thomas Schumacher und Petra Merle für die technische Unterstützung mit dem COPRA-System. Vielen herzlichen Dank an Natalia Maier für die Revision und Korrektur dieser Mammutaufgabe und an Sami Habib für die Hilfe bei der Gestaltung der Präsentation. Besten Dank an Frau Jung für die ratsamen Wörter und Dr. Mathias Schwabe für die Ideen. Danke auch an allen die mich in Stich gelassen haben.

Ehrenwörtliche Erklärung

Ich versichere, dass ich die Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Ich habe die Masterarbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mainz, den 18. November 2022

Abel Hodelín Hernández

Erklärung zu Eigentum und Urheberrecht

Hiermit erkläre ich mein Einverständnis, dass die Hochschule Mannheim, die vorliegende Masterarbeit den Studierenden und interessierten Dritten zur Einsichtnahme zur Verfügung stellt und unter Nennung meines Namens (Urheber) veröffentlichen darf.

Mainz, den 18. November 2022

Abel Hodelín Hernández