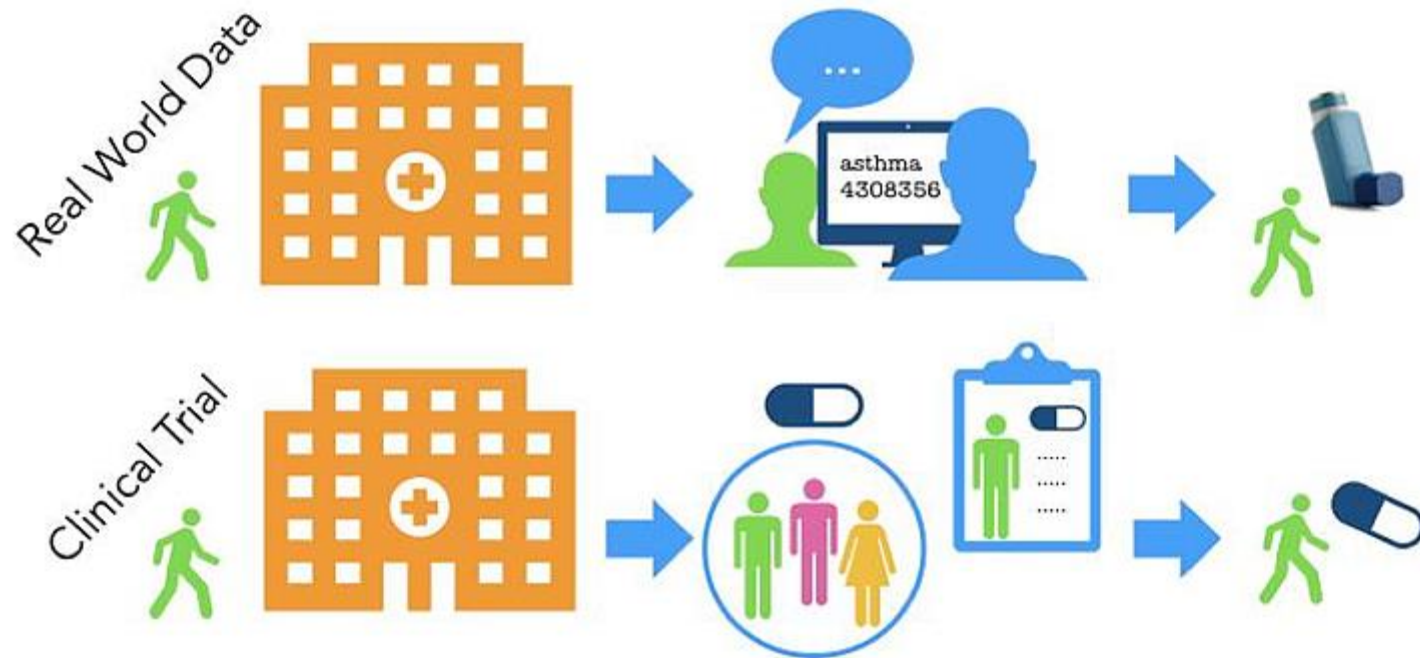


# MDR-gesteuerte Datenqualitäts-Analysen im MIRACUM Konsortium

Dr. med. Lorenz Kapsner

27. März 2020

MIRACUM Kolloquium



[1] Bild aus: <http://blog.thehyve.nl/blog/leveraging-the-omop-common-data-model-for-clinical-trials> (last access 29.08.2019, 12:45)

- Zweck zur Wiederverwendung unterscheidet sich von ursprünglichem Zweck zur Datenerhebung
- Erzeugung der Daten in Systemen, die nicht primär für wissenschaftliche Zwecke geschaffen wurden

**Sorge: Nicht systematisch gesammelte Daten könnten von „schlechter Qualität“ sein**

- **Mögliche Verzerrung von Forschungsergebnisse**

- Im Projektplan verankert
- Entwicklung von Software für Datenqualitätsanalysen (DQ-Analysen) an 2 Standorten:
  - Erlangen
  - Dresden
- An allen Standorten: Erstellung von DQ-Berichten und Implementierung von Feedbackschleifen

# MIRACUM DQA Tool



MIRACUM DQA Tool

Dashboard

Descriptive Results

Plausibility Checks

Atemporal Plausibility

Uniqueness Plausibility

Completeness

Reporting

Config

DQ MDR

Logfile

Reset DQA Tool

MIRACUM DQA Tool

Welcome to your Data-Quality-Analysis Dashboard

Started: 2020-03-17 13:01:59 CET

Finished: 2020-03-17 13:02:36 CET

Duration: 0.62 min.

Completeness Checks (Validation):

Completeness checks (validation) evaluate the 'correctness' of the ETL (extract transform load) jobs. They compare for each variable the exact matching of the number of distinct values, the number of valid values, and the number of missing values between the source data system and the target data system.

Variable	Check Distincts	Check Valids	Check Missings
Alter (in Jahren)	passed	passed	passed
Alter (in Tagen)	passed	passed	passed
Aufnahmeanlass	passed	passed	passed
Aufnahmedatum	passed	passed	passed
Aufnahmedatum (Fachabteilung)	passed	passed	passed
Aufnahmegrund	passed	passed	passed
Beatmungsstunden	passed	passed	passed
Diagnoseart	passed	passed	passed
Diagnosen (ICD)	failed	passed	passed

Value Conformance Checks (Verification):

Value conformance checks (verification) compare for each variable the values of each data system to predefined constraints. Those constraints can be defined for each variable and data system individually in the metadata repository (MDR).

Variable	Check Source Data	Check Target Data
Alter (in Jahren)	passed	passed
Alter (in Tagen)	passed	passed
Aufnahmeanlass	passed	passed
Aufnahmegrund	passed	passed
Beatmungsstunden	passed	passed

Target System Overview (Data Map)

System name: i2b2

Variable	# n	# Valid	# Missing	# Distinct
Medikation (OPS)	3191	3191	0	466
Patientennummer	57436	57436	0	57436
Fallnummer	86876	86876	0	86876
Prozeduren (OPS)	454226	454226	0	8378
Diagnosen (ICD)	479145	479145	0	7180

Send Datamap

Source System Overview

System name: p21csv

Variable	# n	# Valid	# Missing	# Distinct
Medikation (OPS)	3191	3191	0	466
Patientennummer	57436	57436	0	57436
Fallnummer	86876	86876	0	86876
Prozeduren (OPS)	454226	454226	0	8378
Diagnosen (ICD)	479145	479145	0	7673

Version:

DQAstats: 0.1.0

DQAgui: 0.1.0

MIRACUMDQA: 2.0.5

© Universitätsklinikum Erlangen

- Interaktive Webanwendung
- Anbindung an zentrales Metadaten Repository (MDR)
  - Flexible Erweiterbarkeit durch Abbildung neuer Datenelemente im MDR
  - Automatisierte Auswertung basierend auf Datentyp
  - Automatisierte Auswertung der DQ-Dimensionen
  - Parametrierbar auf Datenquellen (§21-CSV, i2b2, OMOP, ...)
  - Kein „hard-coding“ von Datenelementen im Quellcode

## EDM Forum EDM Forum Community

eGEMs (Generating Evidence & Methods to  
improve patient outcomes)

Publish

9-11-2016

## A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Michael G. Kahn

*University of Colorado Anschutz Medical Campus, michael.kahn@ucdenver.edu*

Tiffany J. Callahan

*University of Colorado Anschutz Medical Campus, tiffany.callahan@ucdenver.edu*

Juliana Barnard

*University of Colorado Anschutz Medical Campus, juliana.barnard@ucdenver.edu*

Alan E. Bauck

*Kaiser Permanente Northwest, alan.bauck@kpchr.org*

*See next pages for additional authors*

Framework, bestehend aus

DQ Kategorie	Sub-Kategorien
<b>1. Conformance</b> (Übereinstimmung)	Value conformance
	Relational conformance
	Computational conformance
<b>2. Completeness</b> (Vollständigkeit)	
<b>3. Plausibility</b> (Plausibilität)	Uniqueness plausibility
	Atemporal plausibility
	Temporal plausibility



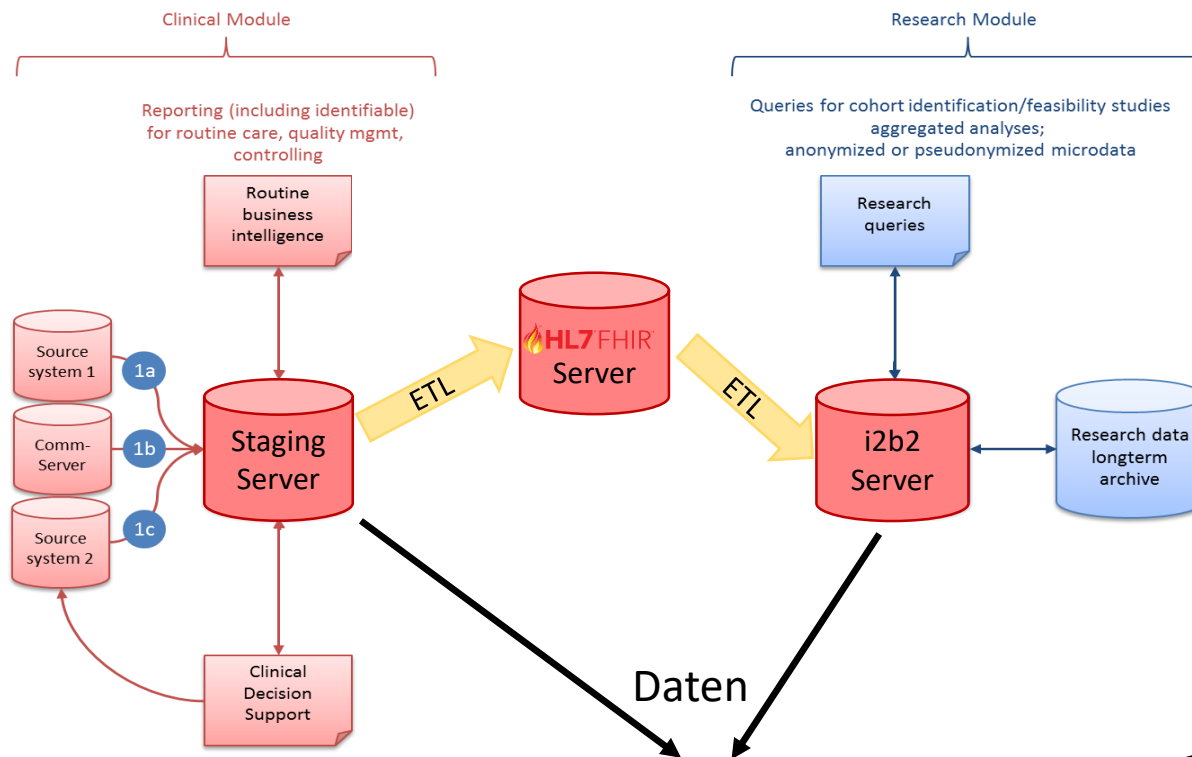
- **Übereinstimmung:** Stimmen Daten mit internen oder externen Vorgaben überein? (z.B. *Data dictionary*)
- **Vollständigkeit:** Sind Daten vollständig vorhanden?
- **Plausibilität:** Sind Daten glaubwürdig und “richtig”?
  - *Uniqueness plausibility:* Multiples Auftreten von Werten in Zusammenhängen, in welchen keine Duplikate vorkommen sollten
  - *Atemporal plausibility:* Überprüfung von Beziehungen zwischen Variablen, deren erwarteter Zusammenhang bekannt ist

- „Uniqueness Plausibility“

Plausibilitätsprüfung
Mit jeder Fallnummer darf nur eine Hauptdiagnose assoziiert sein.

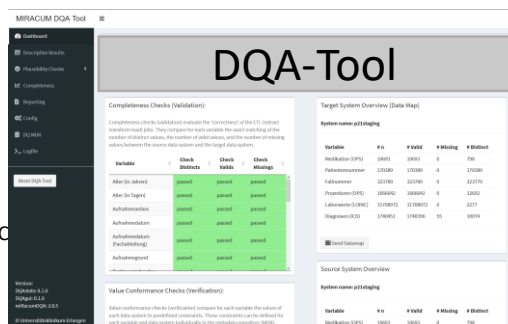
- „Atemporal Plausibility“

Plausibilitätsprüfung	Erwartungswert
Nur bei weiblichen Patientinnen ist eine ICD-Diagnose aus dem ICD-Kapitel XV (ICD O00-O99) (Schwangerschaft, Geburt und Wochenbett) als Krankenhausdiagnose erlaubt	Geschlecht ist “weiblich”.

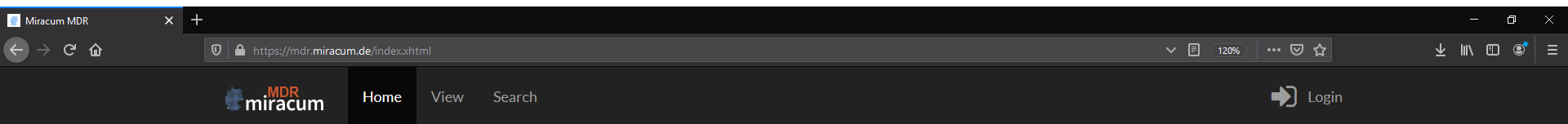


- Datenelemente
- Datenqualitäts-Checks

[1] H.-U. Prokosch et al., MIRACUM: Medical Informatics in Research and C Inf. Med. 57 (2018) e82–e91. doi:10.3414/ME17-02-0025.



# MIRACUM MDR (Samplly.MDR)



## The Metadata Repository (MDR)

The Miracum Metadata Repository serves the harmonization of data content. For this purpose it contains a collection of metadata (definitions of data elements), which the registry compounds record in addition to the minimal data set for rare diseases, which is still contained. The more registries define their data elements in a comprehensible way and use identical data definitions, the easier it is to use the data for overarching questions. In defining the data elements, it is important that their meaning is described accurately and that existing data elements are used if possible.

Please be careful that you do not enter contents that break current law or third-party rights and be aware that the contents are visible and reusable for all registered users!

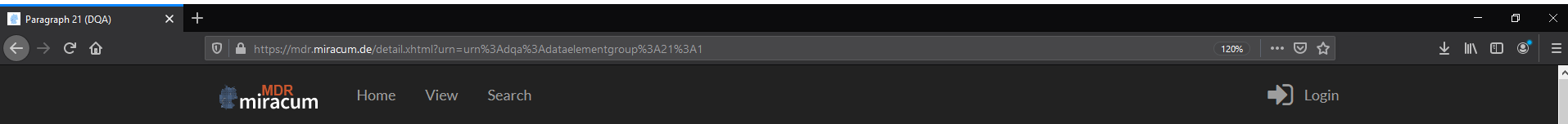


# miracum

Medical Informatics in Research and Care in University Medicine



# MIRACUM MDR (Samplify.MDR)



## Paragraph 21 (DQA)

Diese Gruppe enthält die Datenelemente des Paragraph 21, die im Rahmen der DQ-Analyse geprüft werden.

Released

Revisions ▾



urn:dqa:dataelementgroup:21:1

### Members and subgroups

Alter (in Jahren)

urn:dqa:dataelement:378:1

Die Altersangabe wird für die DRG-Zuordnung benötigt.

Validationtype

Integer

Range

$1 \leq x \leq 115$

Unit of measure

a



### Slots

Name

Value

```
{
  "variable_name": "encounter_subject_patient_age_years",
  "fhir": "undefined",
  "csv": "p21csv",
  "base": "dqa_assessment",
  "filter": "1",
  "key": "dt.ageinyears",
  "variable_name": "encounter_subject_patient_age_years",
  "fhir": "undefined",
  "source_variable_name": "ALTER IN LAUBEN AM AUENALIMETAC"
}
```

## Slots

Name

Value

dqa

- DQ-Logik
- Maschinenlesbar (JSON)
- Zusatzinformationen/ Mappings für zu analysierende Datensysteme
  - i2b2, OMOP, Staging-DB, ...
- Plausibilitäts-Checks
- Darstellungsoptionen (Data map)

# Benutzeroberfläche

WWW



MIRACUM DQA Tool

- Dashboard
- Descriptive Results
- Plausibility Checks
  - Atemporal Plausibility
  - Uniqueness Plausibility
- Completeness
- Reporting
- Config
- DQ MDR
- Logfile

Reset DQA Tool

Select variable

Select variable

- Alter (in Jahren)
- Alter (in Tagen)
- Aufnahmeanlass
- Aufnahmedatum
- Aufnahmedatum (Fachabteilung)
- Aufnahmegrund
- Beatmungstunden
- Diagnoseart
- Diagnosen (ICD)
- Entlassungsdatum

Description

Die Altersangabe wird für die DRG-Zuordnung benötigt.

**Beschreibung des Datenelements**

**Liste der Datenelemente**

**Erwarteter Wertebereich (Value Conformance)**

**Quellsystem**

**Metadaten**

Variable name: ALTER\_IN\_JAHREN\_AM\_AUFNAHMETAG

Source table: **FALL.CSV**

FHIR ressource: undefined

DQ-internal Variable Name: encounter\_subject\_patient\_age\_years

Variable type: integer

**Results**

Mean	50.79
Minimum	1
Median	55
Maximum	100
SD	23.84
Negative	0
Zero	0
Positive	83051
OutLo	0
OutHi	0
Skewness	-0.43
Kurtosis	-0.83
Variance	568.27
Range	99

**Completeness Overview**

n:	86876
Valid values:	83051
Missing values:	3825
Distinct values:	100

**Constraining values/rules:**

\$min  
[1] 1

\$max  
[1] 115

\$unit  
[1] "a"

**Value conformance:**  
Conformance check: passed

**Zielsystem**

**Metadaten**

Variable name: nval\_num

Source table: **observation\_fact**

FHIR ressource: undefined

DQ-internal Variable Name: encounter\_subject\_patient\_age\_years

Variable type: integer

**Results**

Mean	50.79
Minimum	1
Median	55
Maximum	100
SD	23.84
Negative	0
Zero	0
Positive	83051
OutLo	0
OutHi	0
Skewness	-0.43
Kurtosis	-0.83
Variance	568.27
Range	99

**Completeness Overview**

n:	86876
Valid values:	83051
Missing values:	3825
Distinct values:	100

**Constraining values/rules:**

\$min  
[1] 1

\$max  
[1] 115

\$unit  
[1] "a"

**Value conformance:**  
Conformance check: passed

Target SQL





# MIRACUM Datamap

<https://datamap.miracum.org>



MIRACUM DQA Tool

Dashboard

Descriptive Results

Plausibility Checks

Attemporal Plausibility

Uniqueness Plausibility

Completeness

Reporting

Config

DQ MDR

Logfile

Reset DQA Tool

Version:

DQAstats: 0.1.0

DQAgui: 0.1.0

miracumDQA: 2.0.5

© Universitätsklinikum Erlangen

WELCOME TO YOUR DATA-QUALITY-ANALYSIS DASHBOARD

Started: 2020-03-17 13:01:59 CET

Finished: 2020-03-17 13:02:36 CET

Duration: 0.62 min.

Completeness Checks (Validation):

Completeness checks (validation) evaluate the 'correctness' of the ETL (extract transform load) jobs. They compare for each variable the exact matching of the number of distinct values, the number of valid values, and the number of missing values between the source data system and the target data system.

Variable	Check Distincts	Check Valids	Check Missings
Alter (in Jahren)	passed	passed	passed
Alter (in Tagen)	passed	passed	passed
Aufnahmearlass	passed	passed	passed
Aufnahmedatum	passed	passed	passed
Aufnahmedatum (Fachabteilung)	passed	passed	passed
Aufnahmegrund	passed	passed	passed
Beatmungsstunden	passed	passed	passed
Diagnoseart	passed	passed	passed
Diagnosen (ICD)	failed	passed	passed

Value Conformance Checks (Verification):

Value conformance checks (verification) compare for each variable the values of each data system to predefined constraints. Those constraints can be defined for each variable and data system individually in the metadata repository (MDR).

Variable	Check Source Data	Check Target Data
Alter (in Jahren)	passed	passed
Alter (in Tagen)	passed	passed
Aufnahmearlass	passed	passed
Aufnahmegrund	passed	passed
Beatmungsstunden	passed	passed

Source System Overview

System name: p21csv

Variable	# n	# Valid	# Missing	# Distinct
Medikation (OPS)	3191	3191	0	466
Patientennummer	57436	57436	0	57436
Fallnummer	86876	86876	0	86876
Prozeduren (OPS)	454226	454226	0	8378
Diagnosen (ICD)	479145	479145	0	7673

MIRACUM Research Portal / Datamap

909.690 Patienten

1.846.510 Fälle

Laborwerte 80.124.492

Diagnosen 9.207.021

Prozeduren 7.946.597

Medikation (OPS) 44.789

© 2020 MIRACUM Consortium

1.10.0

# Vielen Dank



## Team

- Jonathan Mang
- Susanne Seuchter
- Noemi Deppenwiese
- Sebastian Mate
- Abishaa Vengadeswaran
- Franziska Bathelt
- Mirko Gruhl
- Detlef Kraska
- Ulli Prokosch



Dr. med. Lorenz Kapsner  
[lorenz.kapsner@uk-erlangen.de](mailto:lorenz.kapsner@uk-erlangen.de)



MITGLIED DER



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung  
FKZ 01ZZ1801A



FRIEDRICH-ALEXANDER  
UNIVERSITÄT  
ERLANGEN-NÜRNBERG  
MEDIZINISCHE FAKULTÄT

Universitätsklinikum  
Erlangen

