Patientenzentriertes Dashboard, Transformation longitudinaler Daten und Pseudonymisierung: Lösungen aus dem DIZ der Charité

MIRACUM-DIFUTURE School 2024

Fabian Prasser und Team

24.09.2024

CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN

BIH Berlin Institute of Health @Charité

Agenda

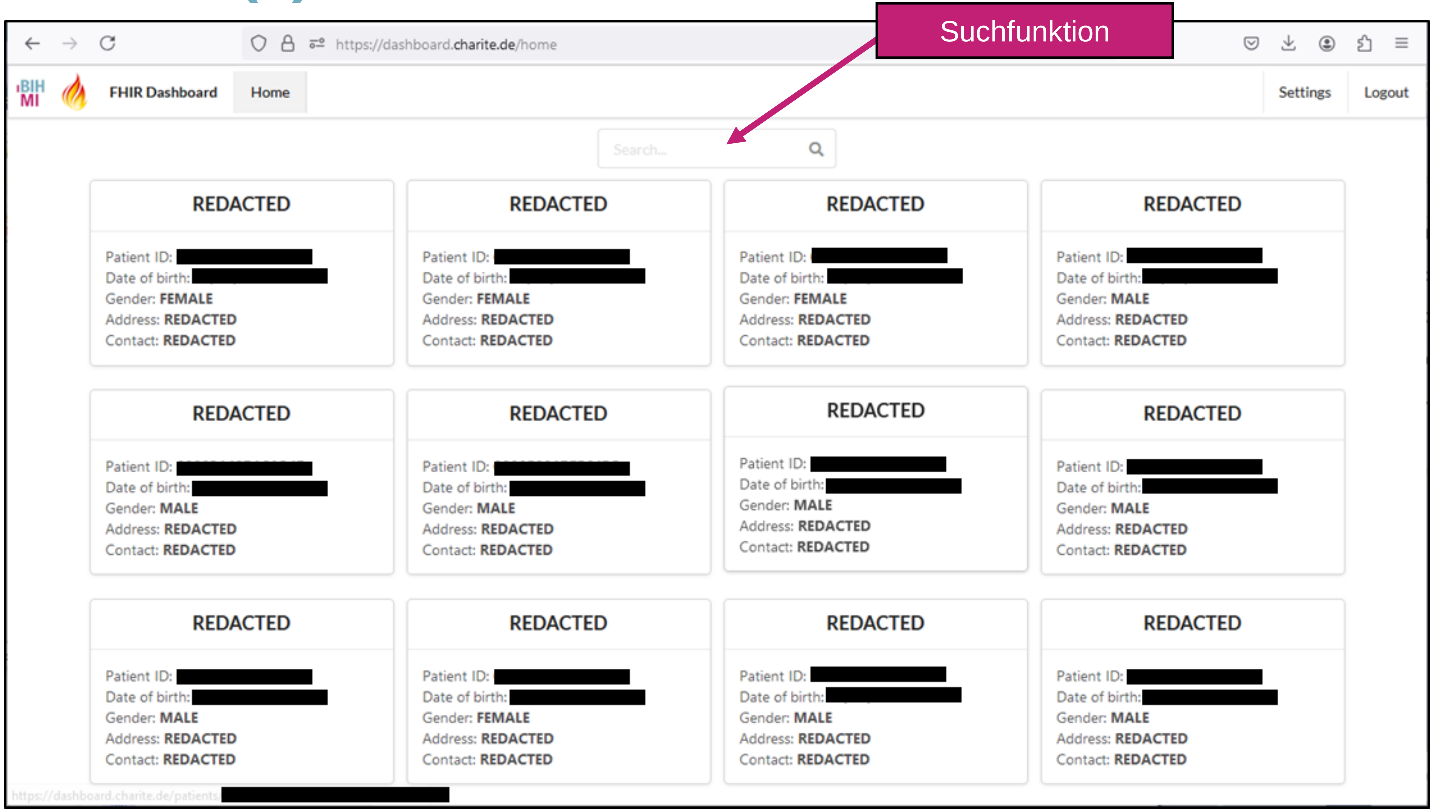
1. Patientenzentriertes FHIR-Dashboard
2. Transformation longitudinaler Daten mit HERALD
3. Pseudonymisierung mit ACE

1. Patientenzentriertes FHIR-Dashboard

Hintergrund

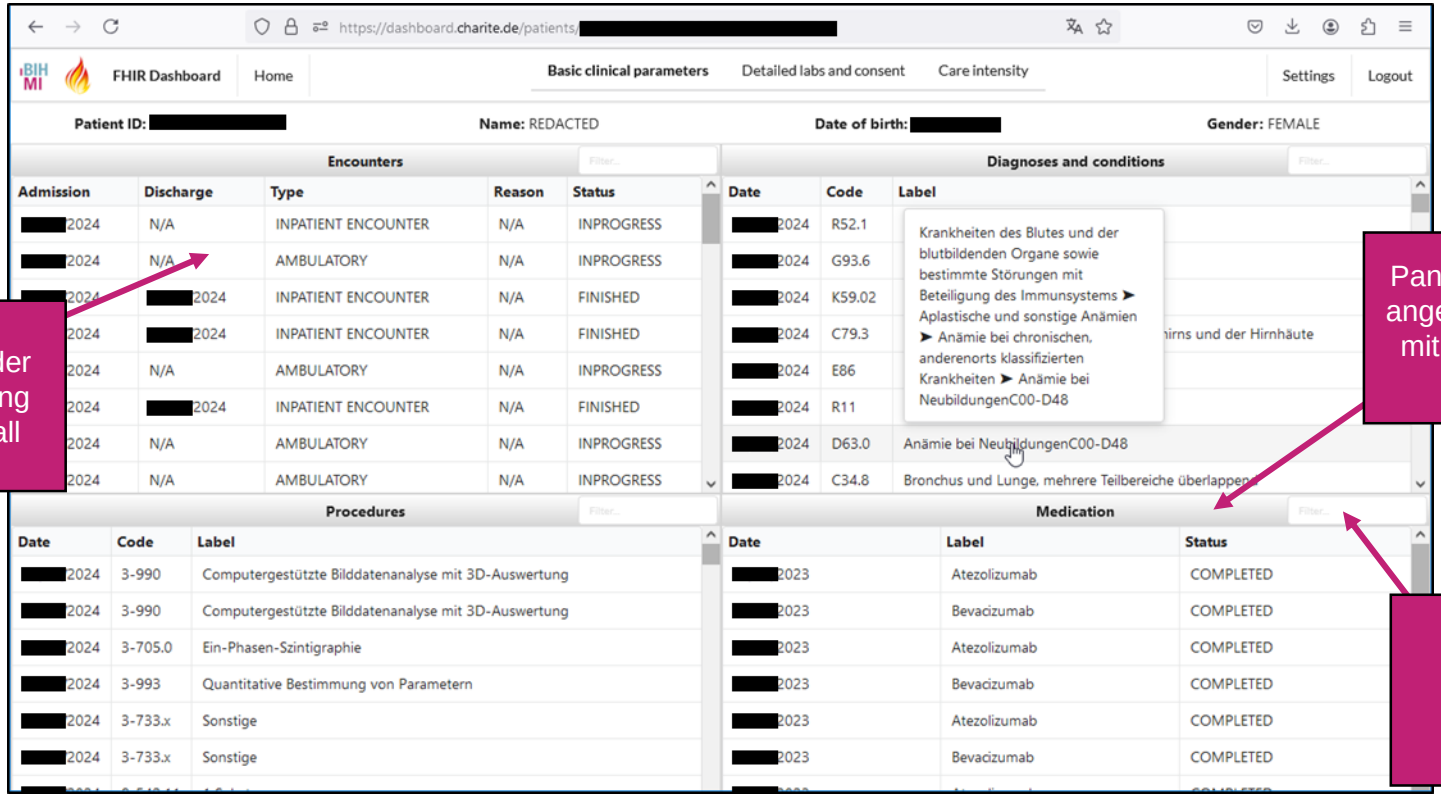
- Verfügbarkeit von FHIR in den DIZen
- Darstellung als patientenzentrierte „Read-Only-Akte“
- Macht das DIZ und den KDS „anfassbar“
- Hilft beim debuggen der FHIR-Daten bzw. Transformationsprozesse
- Perspektivisch Potenzial für klinischen Nutzen, bspw. Befundung
- Integration weitergehender Funktionen, wie Terminologien, und Integration unterschiedlicher Quellsysteme, bspw. für PROMs

Screenshots (1)



CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN
BIH Berlin Institute of Health @Charité

Screenshots (2)

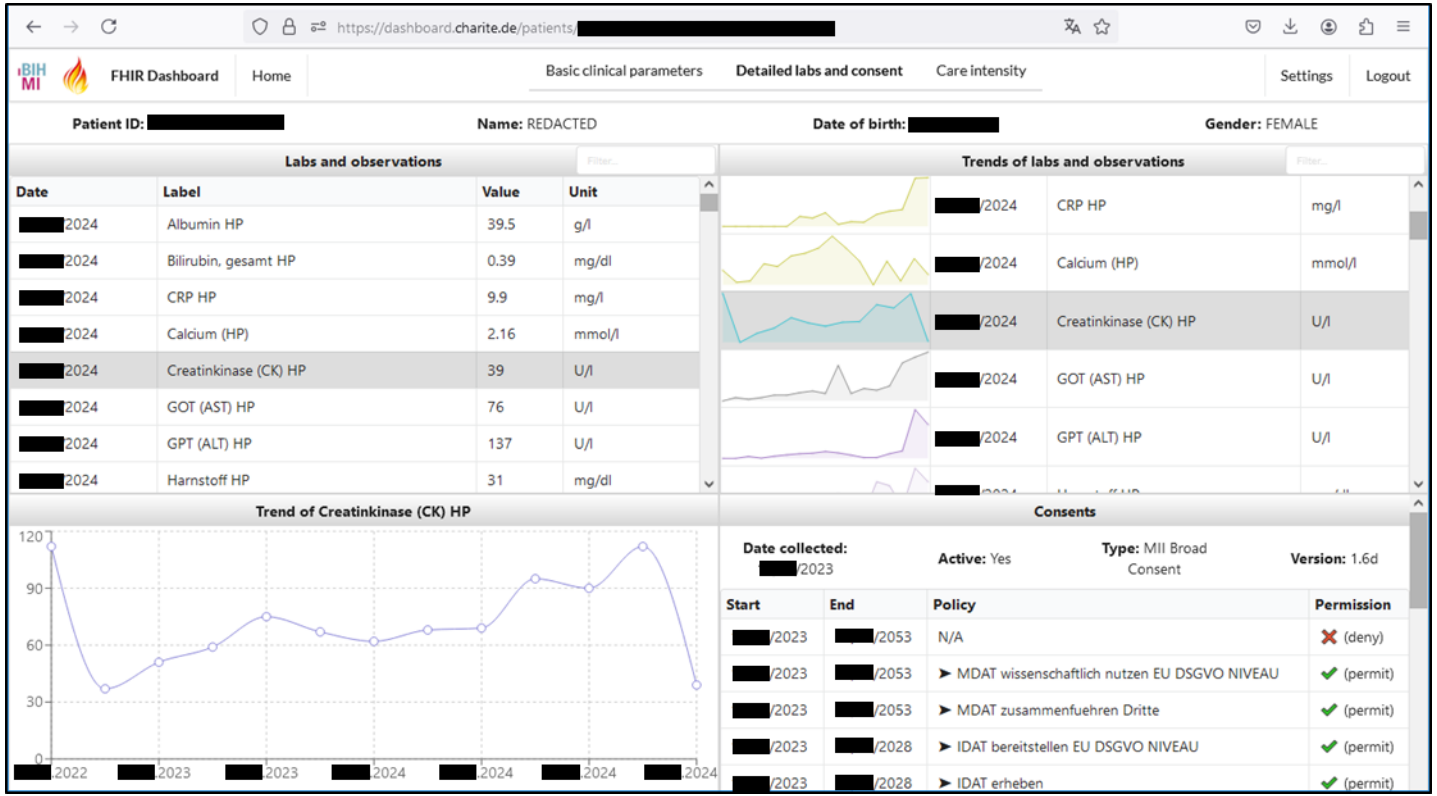


Möglichkeit der Einschränkung auf einen Fall

Panels können frei angeordnet werden mit Persistenz im User-Profil

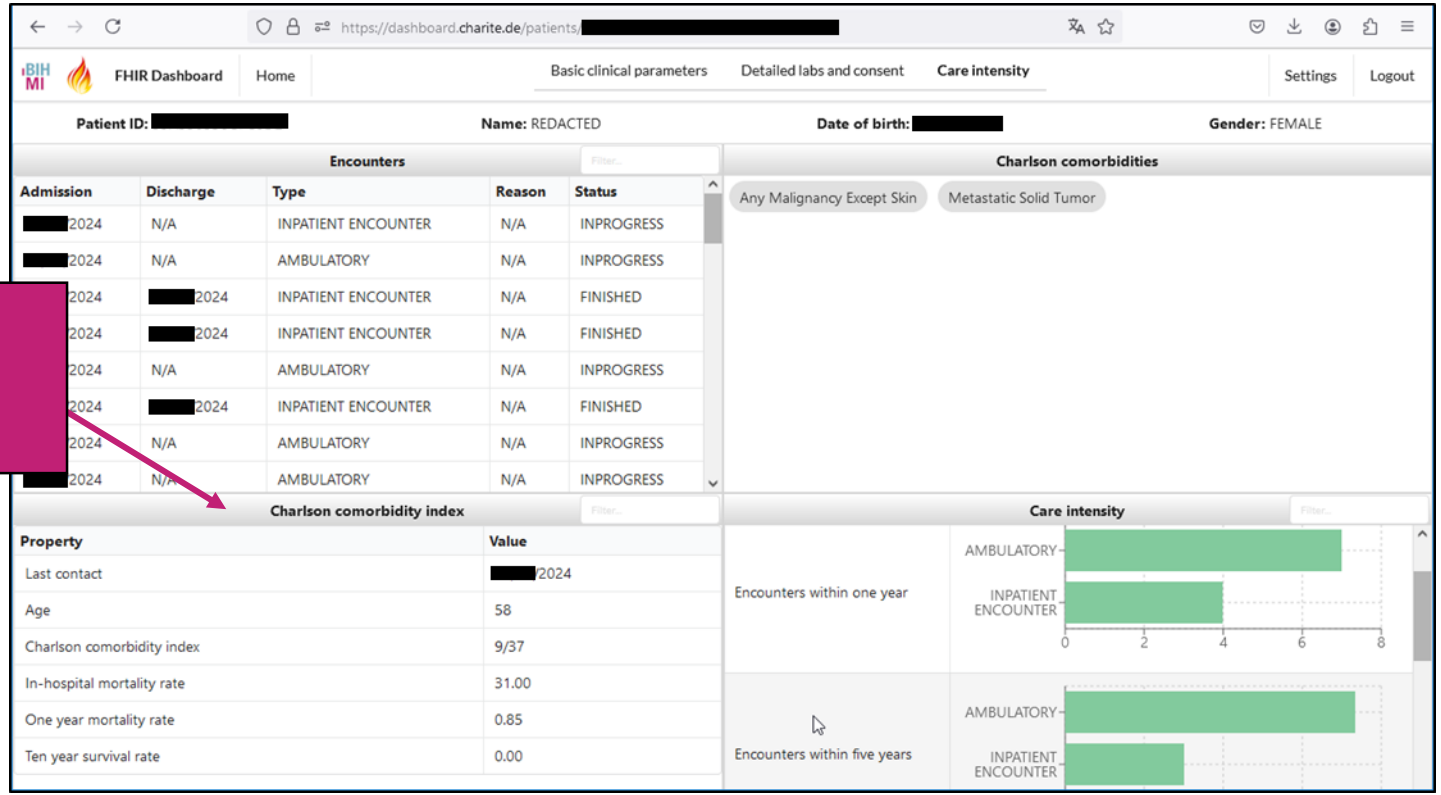
Suchfunktion innerhalb der Panels

Screenshots (3)



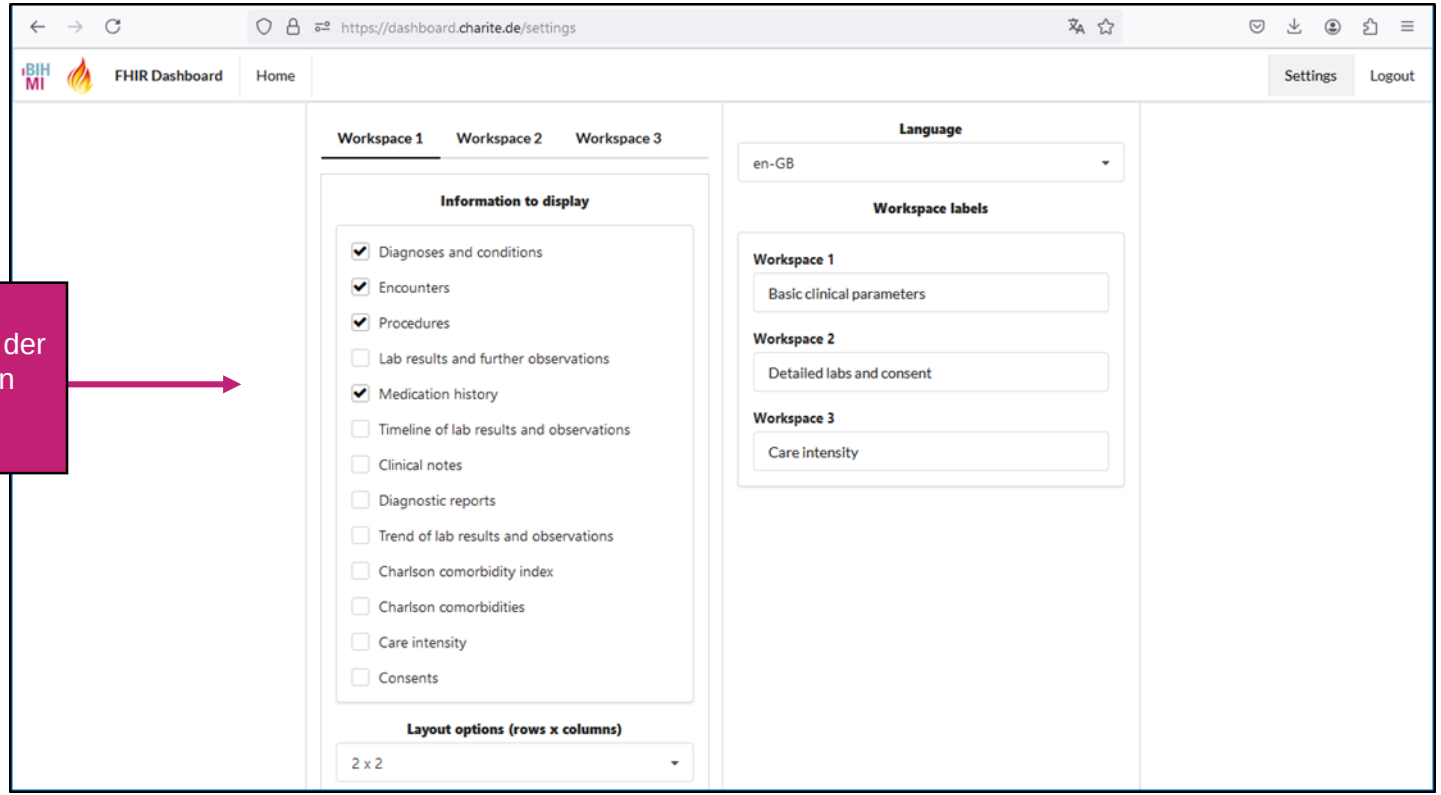
Screenshots (4)

Nur eine Spielerei

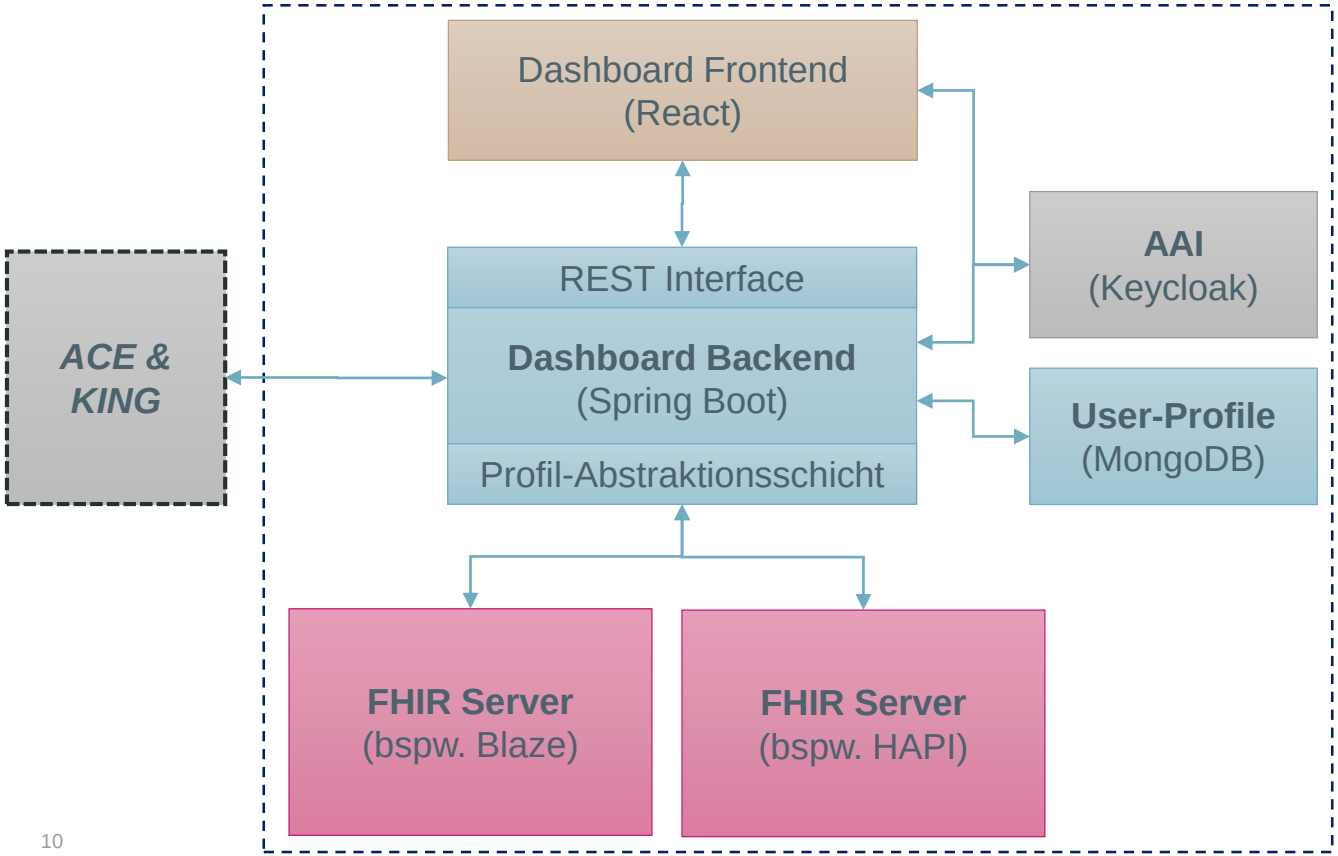


Screenshots (5)

Konfiguration der individuellen Ansicht



Architektur und Technologie

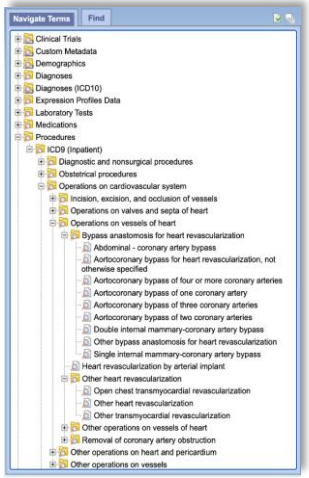


Vollständig oder teilweise deploybar per Docker

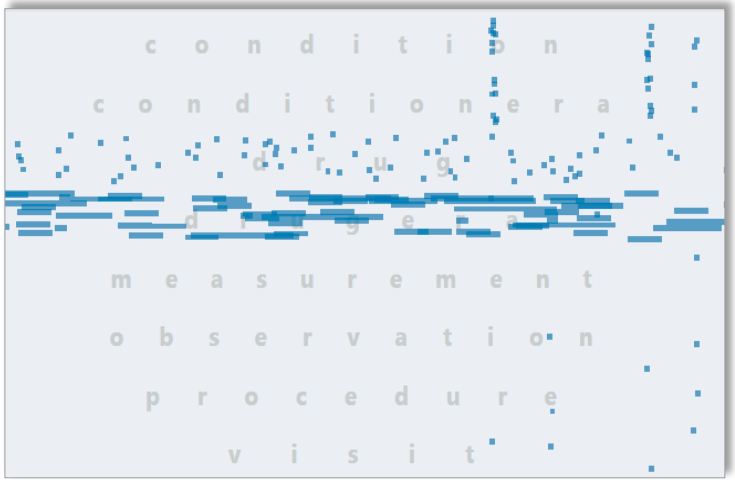
2. Transformation longitudinaler Daten mit HERALD

Herausforderungen

Klinische Routinedaten sind …



…hierarchisch



…longitudinal

Viele Analysen benötigen…

Patient ID (#Values: 338)	Age (#Values: 338)	Sex (#Values: 338)	FirstCancerDiagnosis (#Values: 338)	FirstChemoRadio (#Values: 338)	ProcedureCount (#Values: 338)	Medication (#Values: 328)	Hemoglobin (#Values: 151)	Glucose (#Values: 144)
120820	51	M	Non-small cell carcinoma of lung TNM stage 4 (disorder)	Combined chemotherapy and radiation therapy (procedure)	9	PACLitaxel 100 MG Injection	5.90000 %	69.00000 mg/dL
121087	60	M	Non-small cell carcinoma of lung TNM stage 4 (disorder)	Combined chemotherapy and radiation therapy (procedure)	8	Cisplatin 50 MG Injection		
121125	46	F	Non-small cell carcinoma of lung TNM stage 4 (disorder)	Combined chemotherapy and radiation therapy (procedure)	4	PACLitaxel 100 MG Injection		
121287	46	F	Non-small cell carcinoma of lung TNM stage 4 (disorder)	Combined chemotherapy and radiation therapy (procedure)	7	PACLitaxel 100 MG Injection	6.00000 %	82.00000 mg/dL
121505	44	F	Non-small cell carcinoma of lung TNM stage 4 (disorder)	Combined chemotherapy and radiation therapy (procedure)	8	Cisplatin 50 MG Injection	6.90000 %	117.00000 mg/dL
121934	52	F	Non-small cell carcinoma of lung TNM stage 4 (disorder)	Combined chemotherapy and radiation therapy (procedure)	2	Cisplatin 50 MG Injection	6.30000 %	83.00000 mg/dL

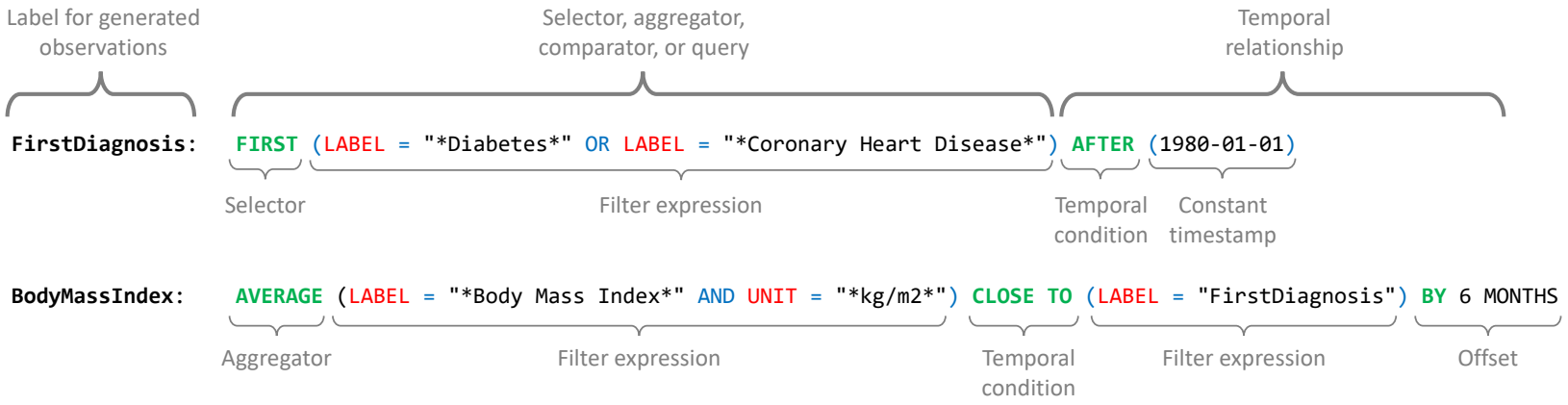
…tabulare Querschnittsdaten

Ansatz: Domänenspezifische Sprache HERALD

A.

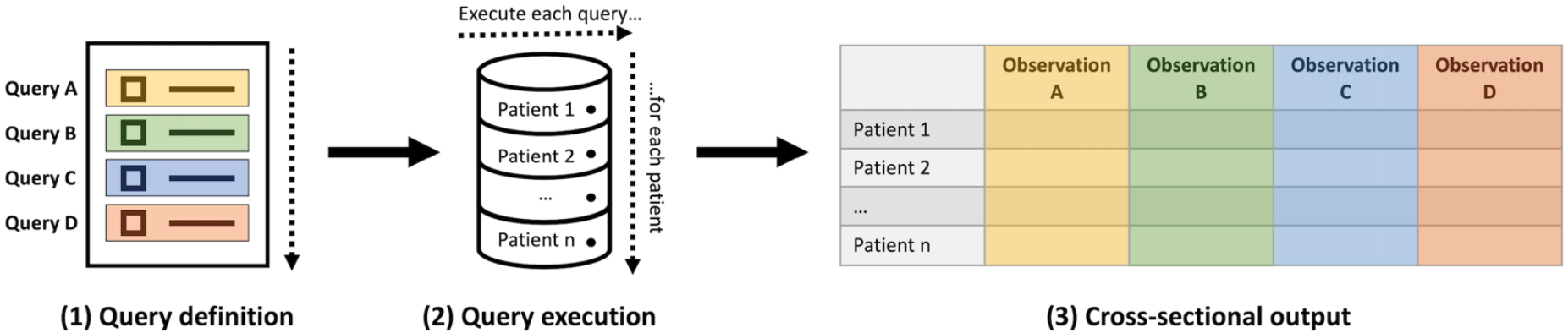
```
FirstDiagnosis:    FIRST (LABEL = "*Diabetes*" OR LABEL = "*Coronary Heart Disease*") AFTER (1980-01-01)
Comorbidities:     EXISTS (LABEL = "*Diabetes*") AND EXISTS (LABEL = "*Coronary Heart Disease*")
BodyMassIndex:     AVERAGE (LABEL = "*Body Mass Index*" AND UNIT = "*kg/m2*") CLOSE TO (LABEL = "FirstDiagnosis") BY 6 MONTHS
LDLCholesterol:    FIRST (LABEL = "*LDL Cholesterol*" AND UNIT = "*mg/dL*") AFTER (LABEL = "FirstDiagnosis") BY 4 WEEKS
HDLCholesterol:    ANY (LABEL = "*HDL Cholesterol*" AND UNIT = "*mg/dL*") CLOSE TO (LABEL = "LDLCholesterol") BY 1 DAY
CholesterolRatio:  RATIO BETWEEN (LABEL = "LDLCholesterol") AND (LABEL = "HDLCholesterol")
```

B.

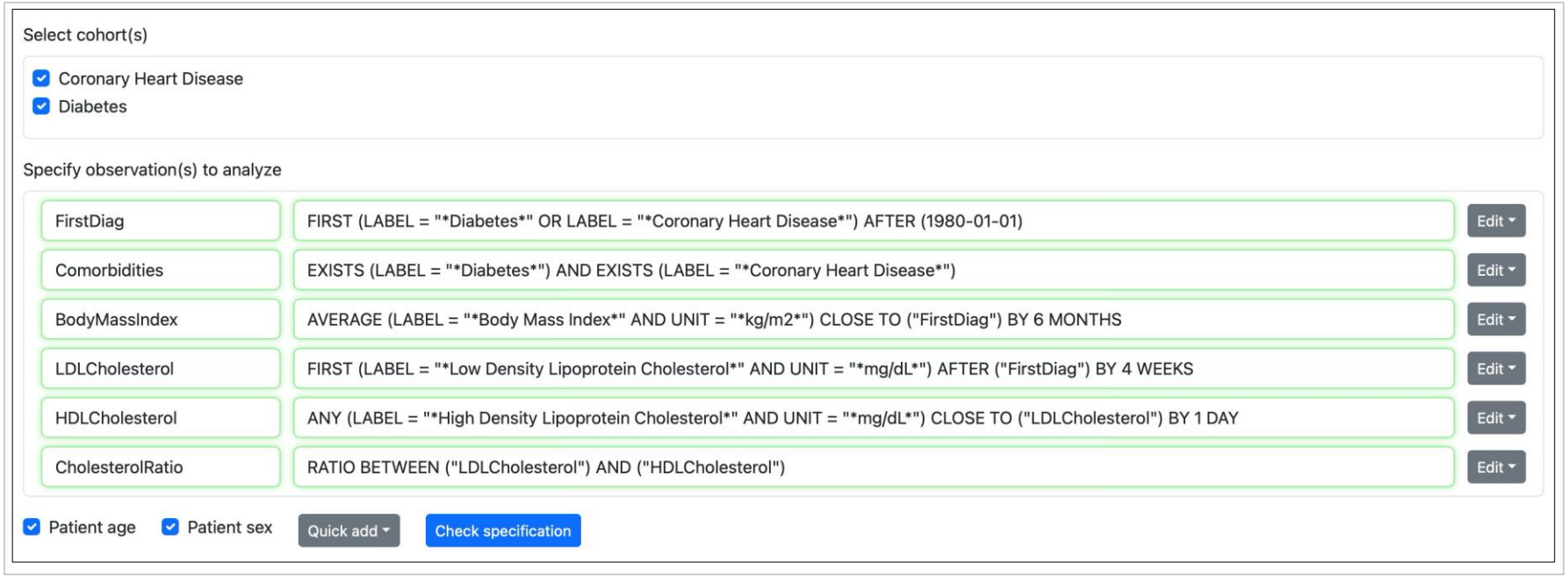


Ausführung von HERALD Queries

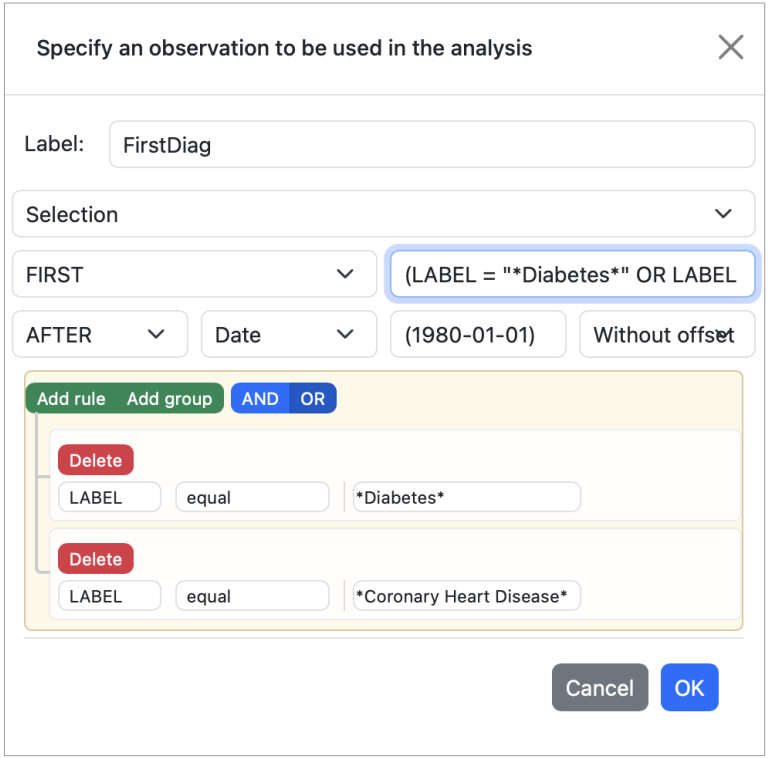
- Eine Datendefinition ist eine Liste von Queries, die jeweils eine Beobachtung produzieren
- Jede Query verarbeitet alle Beobachtungen als Eingabe und erzeugt eine Beobachtung
- Queries werden nacheinander für jeden Patienten in der aufgeführten Reihenfolge ausgeführt
- Queries können sich auf Beobachtungen beziehen, die von früheren Queries erzeugt wurden



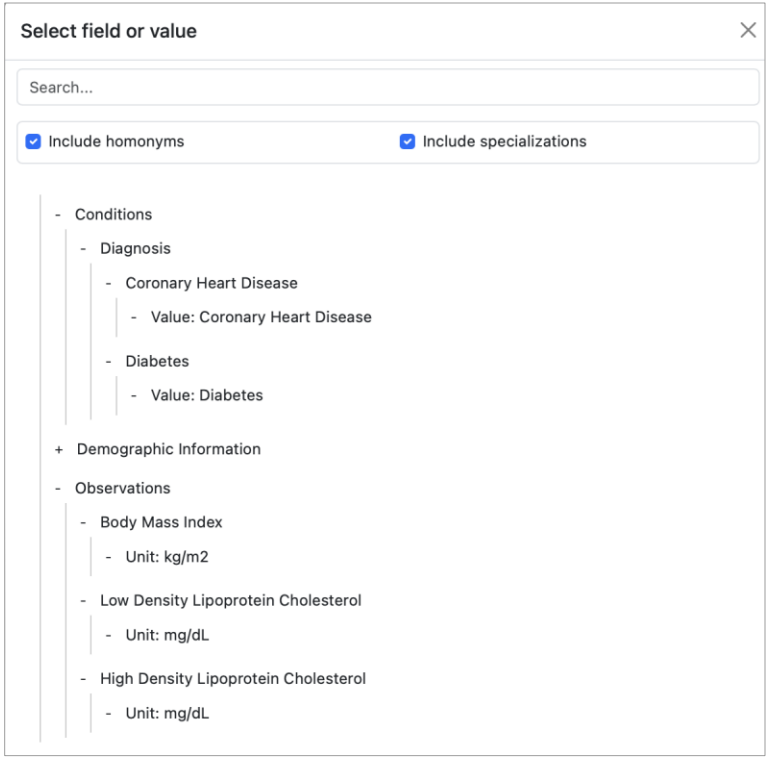
Beispielimplementierung: Datendefinition



Query-Editor



Ontologie-Ansicht



Beispielimplementierung: Analysen und Visualisierungen

A. Data quality report

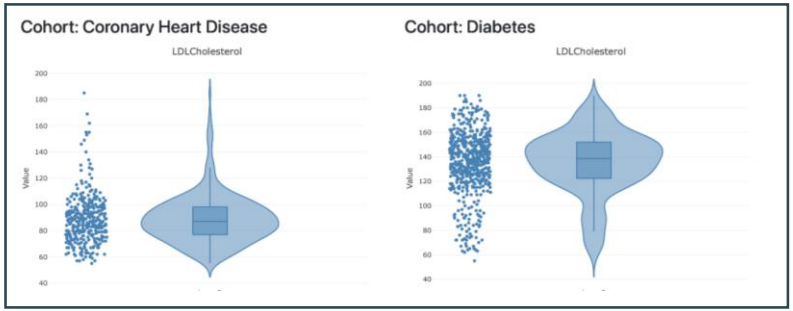
Data quality report for Cohort Coronary Heart Disease

Label	Completeness (%)	Missing rate (%)	Uniqueness (%)	Data Type	Unit
FirstDiag	100.00	0.00	99.56	Categorical	Unknown
Comorbidities	100.00	0.00	100.00	Categorical	Boolean
BodyMassIndex	66.18	33.82	100.00	Decimal	kg/m2
LDLCholesterol	48.25	51.75	99.70	Integer	mg/dL
HDLCholesterol	45.92	54.08	99.68	Integer	mg/dL
CholesterolRatio	45.92	54.08	100.00	Decimal	(mg/dL) / (mg/dL)

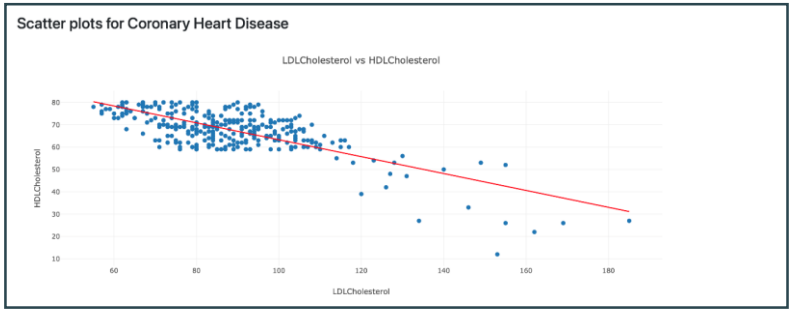
Data quality report for Cohort Diabetes

Label	Completeness (%)	Missing rate (%)	Uniqueness (%)	Data Type	Unit
FirstDiag	100.00	0.00	99.85	Categorical	Unknown
Comorbidities	100.00	0.00	100.00	Categorical	Boolean
BodyMassIndex	76.30	23.70	99.80	Decimal	kg/m2
LDLCholesterol	80.48	19.52	99.81	Integer	mg/dL
HDLCholesterol	78.69	21.31	99.43	Integer	mg/dL
CholesterolRatio	78.69	21.31	100.00	Decimal	(mg/dL) / (mg/dL)

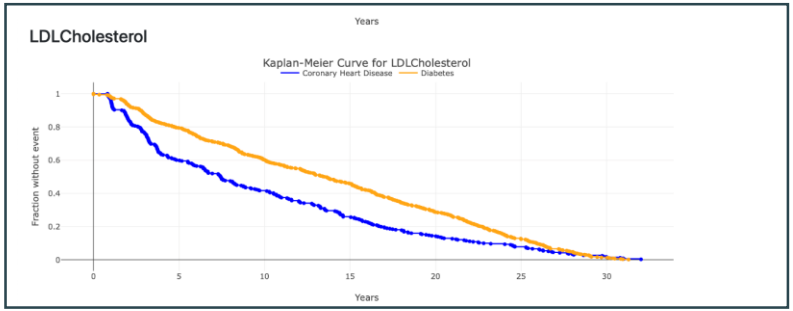
B. Cohort comparison



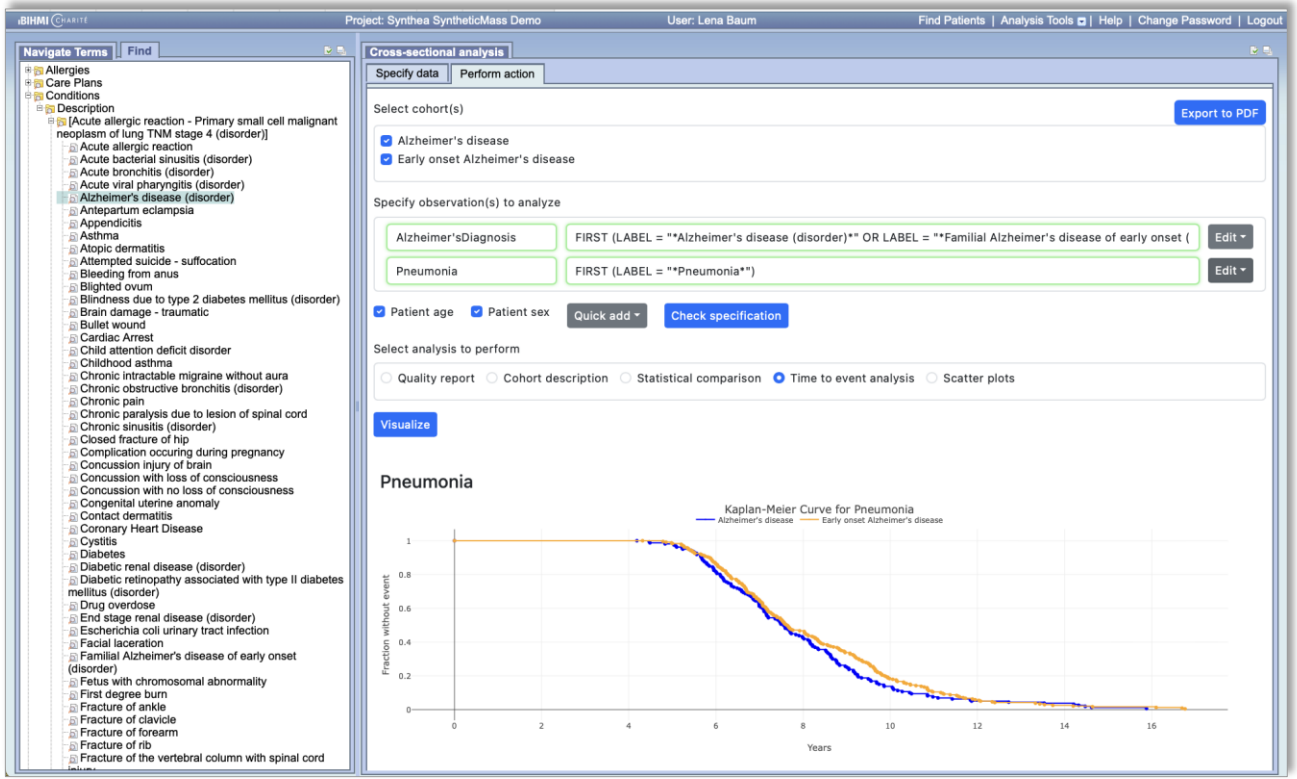
C. Scatter plot



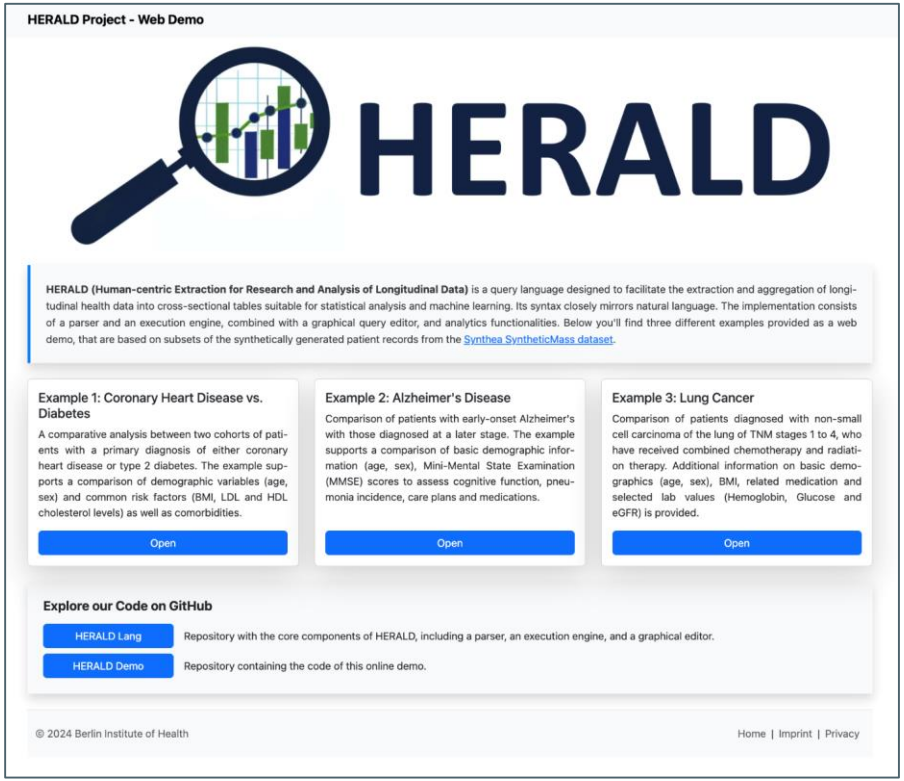
D. Time-to-event analysis



Beispielimplementierung: i2b2-Integration



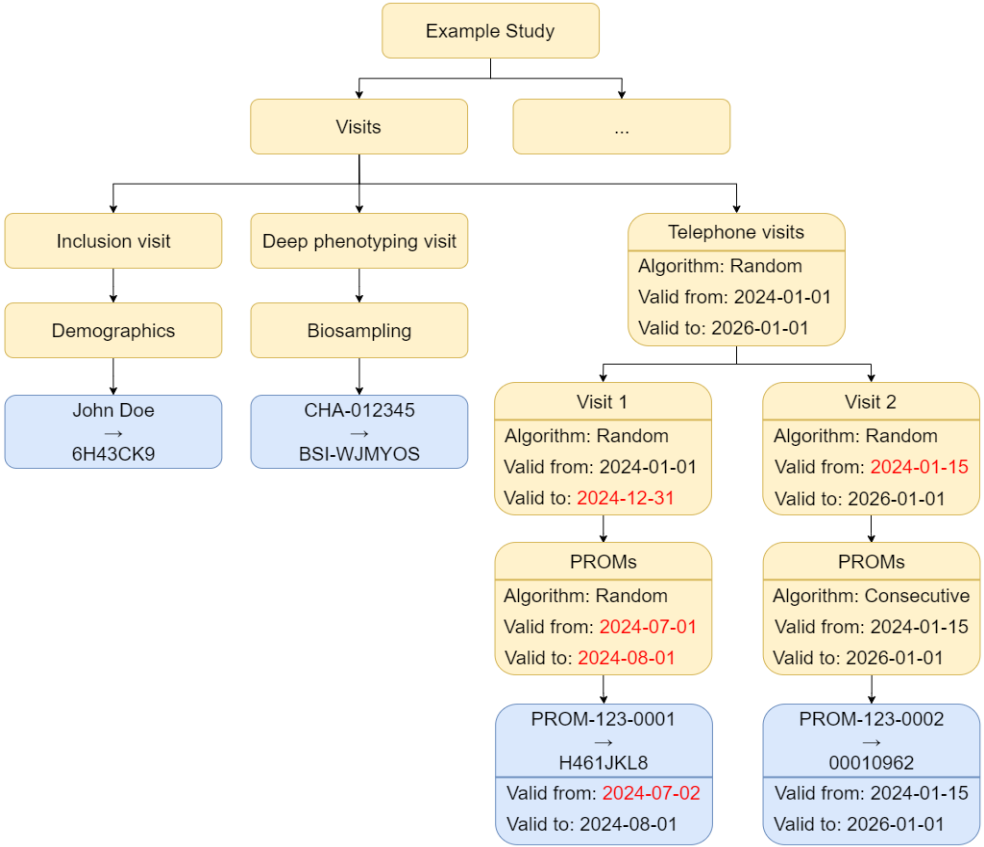
Weiterführende Informationen und Online-Demo



https://www.herald-lang.org

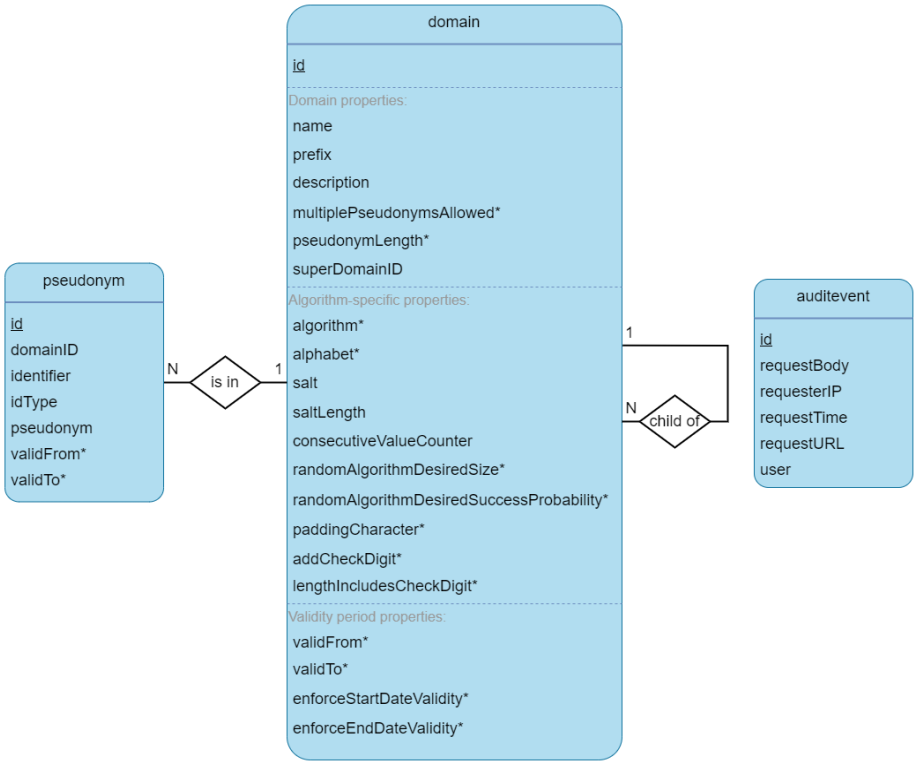
3. Pseudonymisierung mit ACE

Herausforderung und Lösung



- Komplexe Anforderungen an die Pseudonymisierung (u.a. primär, sekundär, Datenarten, Audit Trail, Gültigkeiten)
- Hohe Skalierbarkeit benötigt (bspw. im DIZ der Charité 4 Mio. Patienten, 16 Mio. Fälle)

Design und Features

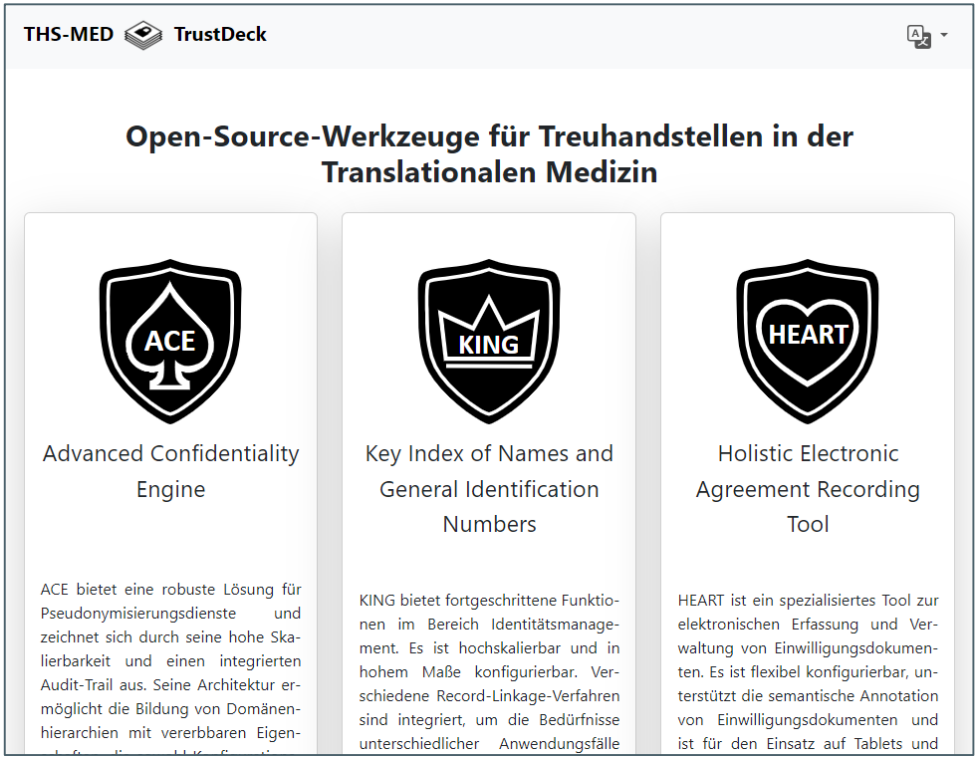


- Hierarchisches Domänenmodell mit Gültigkeitszeiträumen und Vererbung
- Docker, PostgreSQL, Spring Boot, Keycloak
- Berechtigungsmodell: Nutzer *x* Domäne *x* Operation
- Konfigurierbar: Pseudonymisierungsalgorithmus, Alphabet, Checksums
- Vollständiger Audit-Trail

Performance

CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN BIH Berlin Institute of Health @Charité

Weiterführende Informationen



https://www.ths-med.de

Danke für Ihre Aufmerksamkeit!

mi.bihealth.org
medic.charite.de

CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN
BIH Berlin Institute of Health @Charité