

# DATENAUSTAUSCH IN DER FORSCHUNG VIA OMOP/OHDSI

Wenn medizinische Einrichtungen Daten gemeinsam nutzen und für Forschungszwecke gemeinsam auswerten wollen, dann sollten die Einrichtungen die gleiche Sprache sprechen. Sonst wird es unübersichtlich, und die zahlreichen Versprechungen, mit der die Digitalisierung der medizinischen Versorgung antritt, bleiben unerfüllt. Im Rahmen des MIRACUM-Konsortiums nutzen die beteiligten Einrichtungen das aus den USA stammende Datenmodell OMOP und machen damit gute Erfahrungen. Nötig ist in Deutschland in Ermangelung der erforderlichen Terminologien allerdings ein separates Mapping der Routedaten, was (unnötige) Aufwände generieren kann.

TEXT: MARTIN SEDLMAYR, HANS-ULRICH PROKOSCH

**D**ie gemeinsame Nutzung von medizinischen Daten (Data Sharing) steht heute im Mittelpunkt vieler medizinischer Forschungsprojekte. Ein erster Schritt dazu ist die Integration von Daten aus den verschiedenen Quellen innerhalb eines Krankenhauses – beispielsweise den Arbeitsplatzsystemen der verschiedenen Stationen, den Laborsystemen, den PAC-Systemen oder den Apothekensystemen – in ein Data Warehouse oder eine Forschungsdatenbank. Damit sind die Daten zwar lokal verfügbar, aber nicht notwendigerweise an verschiedenen Standorten in der gleichen Struktur oder Terminologie. Es sind also zusätzliche Aufwände für die Harmonisierung zwischen unterschiedlichen Standorten eines Verbundforschungsnetzwerks notwendig.

Die Nutzung eines gemeinsamen Datenmodells (Common Data Model, CDM) und einer gemeinsamen Terminologie vereinfacht die Zusammenarbeit

zwischen den Institutionen. Ein CDM ermöglicht die einfache Zusammenführung medizinischer Daten in größere Datenbanken („die Daten zur Analyse bringen“) oder die Verteilung von Forschungsanalysen und die Aggregation der Ergebnisse („die Analyse zu den Daten bringen“). Im Laufe der letzten Jahre wurden verschiedene CDMs entwickelt, wie zum Beispiel Sentinel zur Überwachung der FDA-regulierten Medizinprodukte, PCORnet für Patientennetzwerke in den USA und CDISC SDTM für klinische Studien.

### DAS OMOP COMMON DATA MODEL

Ein CDM, welches in den letzten Jahren großen Zuspruch erhalten hat, ist das OMOP (Observational Medical Outcomes Partnership) Common Data Model. Ursprünglich wurde es von einer Public-Private-Partnership entwickelt, um die Nutzung von Beobachtungsdaten im Gesundheitswesen und die Evaluation der Auswirkungen

von Medizinprodukten und anderen therapeutischen Maßnahmen und fördern („Observational Outcome Studies“). Der Zweck des OMOP CDM besteht darin, das Format und den Inhalt von Beobachtungsdaten zu standardisieren, so dass standardisierte Anwendungen, Werkzeuge und Methoden auf sie angewendet werden können. Version 1 wurde im Rahmen eines Pilotprojekts im Jahr 2008 in den USA veröffentlicht. Seither gab es vier große Releases, die den Umfang, die Detailtiefe und damit die möglichen Einsatzszenarien stetig vergrößert haben.

Die Observational Health Data Sciences and Informatics (OHDSI) Community setzt seit 2012 die Arbeiten nach Ende der Projektlaufzeit von OMOP fort ([www.ohdsi.org](http://www.ohdsi.org)). OHDSI will den Schatz von Beobachtungs-Gesundheitsdaten durch groß angelegte Analytik heben. Die Forschungsgemeinschaft ermöglicht aktives Engagement über mehrere Disziplinen hinweg, z. B. klinische Medizin, Bio-

statistik, Informatik, Epidemiologie, Biowissenschaften, und umfasst mehrere Interessengruppen, z. B. Forscher, Patienten, Anbieter, Kostenträger, Produkthersteller, Regulierungsbehörden.

### ZAHLREICHE OPEN-SOURCE-WERKZEUGE ERGÄNZEN DAS DATENMODELL

OHDSI bietet nicht nur eine aktualisierte Version des Datenmodells OMOP v5 (Abbildung 1) und einen zentralen Vokabulardienst Athena, sondern auch Open-Source-Werkzeuge und Methoden für verschiedene Arten der Datenanalyse, z. B. Datenqualität und Datencharakterisierung, medizinische Produktsicherheitsüberwachung, vergleichende Wirksamkeit, Versorgungsqualität und prädiktive Modellierung auf Patientenebene. Fast einhundert aktive Entwickler aus verschiedenen Ländern tragen auf [github.com](https://github.com) zu den Werkzeugen bei.

OMOP verwendet sogenannte standardisierte Vokabulare für die

## Common Data Model

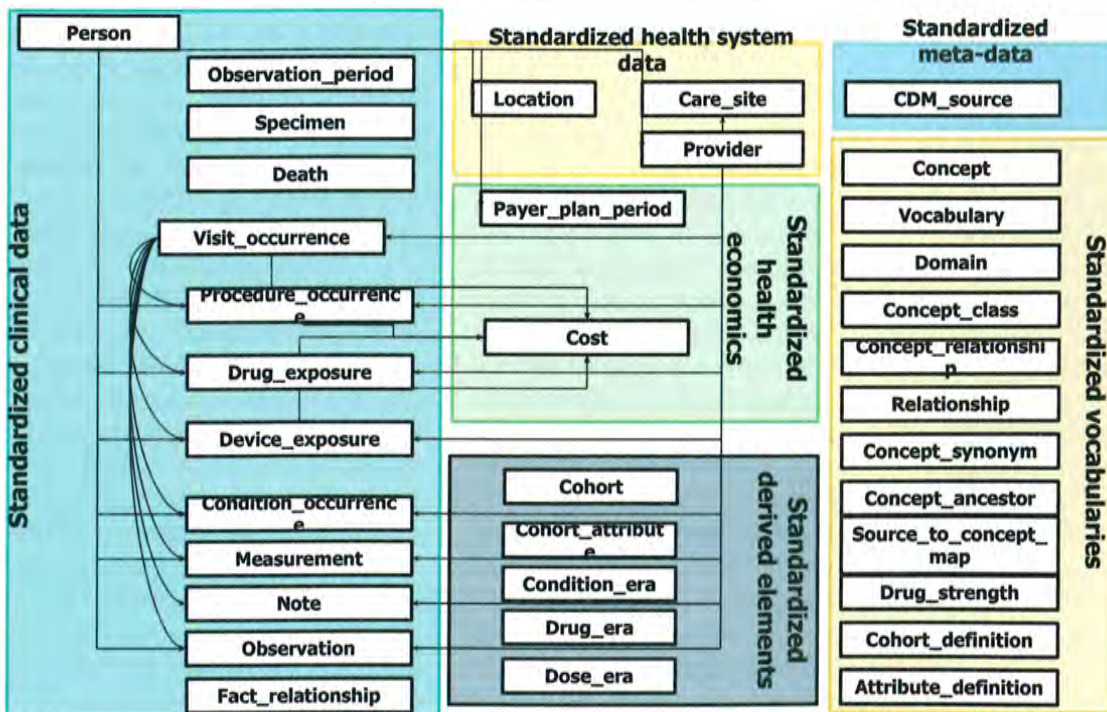
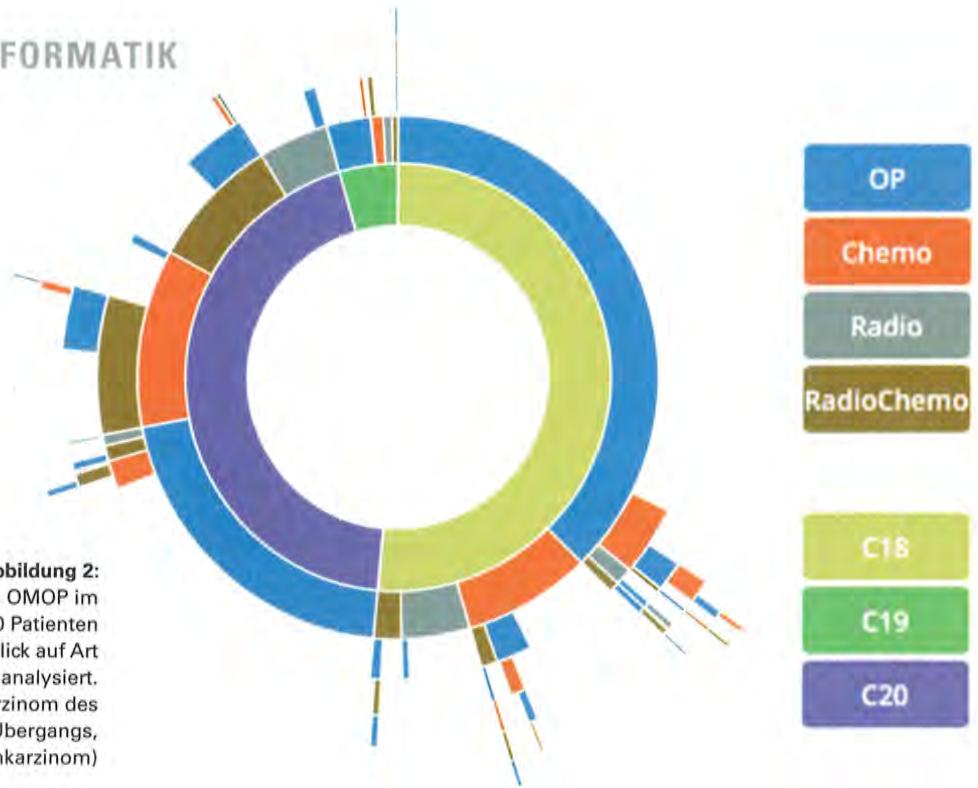


Abbildung 1:  
Das Datenmodell  
OMOP in der derzeit im  
MIRACUM-Konsortium  
genutzten Version 5



**Abbildung 2:**  
 Proof-of-Concept-Studie für den Einsatz von OMOP im Rahmen der Medizininformatik-Initiative: 16 700 Patienten mit kolorektalem Karzinom wurden im Hinblick auf Art der Therapie und Tumorlokalisation analysiert. (C18: Kolonkarzinom, C19: Karzinom des rektosigmoidalen Übergangs, C20: Rektumkarzinom)

## » OMOP/OHDSI wird in mehreren Ländern wie den USA und Korea erfolgreich mit Daten von Hunderten von Millionen von Patienten verwendet. «

Darstellung von Daten im CDM, wie z. B. SNOMED CT für Symptome und Diagnosen oder RxNorm für Medikamente. Mit „Athena“ bietet OHDSI Zugriff auf eine Bibliothek mit mehr als 70 Vokabularen, deren Terme auf die Standardkonzepte abgebildet werden. Der Katalog kann um eigene Terminologien erweitert werden.

Zu den wesentlichen Werkzeugen von OHDSI gehören

- USAGI zur Unterstützung des Terminologiemappings
- WhiteRabbit/RabbitInAHat zur Spezifikation der Abbildung lokaler Datenstrukturen auf das OMOP CDM (Extract Transform Load, ETL-Prozesse)
- Achilles, als quantitative Statistik der Datenbasis mit mehr als 500 Tests zur Prüfung der Datenqualität
- Atlas, eine Webseite für den Forscher zur Suche in den Daten und zur Kohortendefinition
- und mehrere R-Skripte für Patient-level Prediction und Population Estimation, die nicht nur typische Auswertungen auf Kohorten (z. B. beim Vergleich zweier Wirkstoffe) beinhalten, sondern auch die für Publikationen notwendige Definition der Ein- und Ausschlusskriterien in menschenlesbarer Form ausgeben.

OHDSI definiert einen Arbeitsablauf für Forscher, um Forschungsanfragen zu veröffentlichen und an Studien unter OHDSI-Teilnehmern teilzunehmen, die eine aktive Forschungsgemeinschaft fördern. OMOP/OHDSI wird in mehreren Ländern wie den USA und Korea erfolgreich mit Daten von Hunderten von Millionen von Patienten verwendet.

Seit ihrem Start 2012 ist die OHDSI Community stetig gewachsen. Nach den USA sind Südkorea, China und mit dem OHDSI Europe Event im März 2018 in Rotterdam auch Europa als Subcommunities hinzugekommen. Mittlerweile liegen in ca. 80 Institutionen weltweit 1,25 Milliarden Patientenakten (ohne Record Linkage) vor.

### OMOP/OHDSI IM MIRACUM-KONSORTIUM DER MEDIZININFORMATIK-INITIATIVE

In Deutschland werden das OMOP CDM und die OHDSI-Werkzeuge im Projekt MIRACUM (Medical Informatics for Research and Care in University Medicine) eingesetzt. MIRACUM ist eines von vier Konsortien der Medizininformatik-Initiative des BMBF. Es vereint momentan acht Universitätskliniken, zwei Hochschulen und einen Industriepartner aus fünf

deutschen Bundesländern. Ziel ist es, klinische Daten, Bilddaten und Daten aus molekularen/genomischen Untersuchungen sowohl standortbezogen als auch standortübergreifend über modular aufgebaute, skalierbare und föderierte Datenintegrationszentren (DIZ) für innovative Forschungsprojekte nutzbar zu machen.

Im DIZ von MIRACUM werden die Vielfalt der klinischen Daten erschlossen, harmonisiert und dem Forscher zur Verfügung gestellt. Dies erfordert eine Architektur, die die Vielzahl der eingesetzten Komponenten zu einem leistungsfähigen, aber sicheren Ökosystem verbindet. Dieses Ökosystem trägt im Projekt MIRACUM den Namen MIRACOLIX: Medical Informatics Reusable eCO-system of open source Linkable and Interoperable software tools. MIRACUM setzt bei den Komponenten stark auf Open-Source-Werkzeuge, die sich national und international in Forschungsprojekten bewährt haben.

Die Architektur unterscheidet zwei wesentliche Bereiche: Im klinischen Bereich werden die Daten zunächst aus verschiedenen Systemen zusammengeführt und harmonisiert. Im Forschungsbereich werden sie nach erfolgter Deidentifizierung dem Forscher – nach einem Freigabeprozess – zur Verfügung gestellt. Das OMOP CDM ist hierfür ein wichtiger Baustein.

#### **PROOF-OF-CONCEPT UNTER DEUTSCHEN BEDINGUNGEN**

Während die OHDSI Community bereits mehr als 300 Publikationen und eine Vielzahl von Studien durchgeführt hat, wurde das OMOP CDM in Deutschland bisher noch nicht eingesetzt. Daher wurde in der Konzeptphase der Medizininformatik-Initiative ein Proof-of-Concept durchgeführt, um die Möglichkeiten von OMOP zur Nutzung in Deutschland zu evaluieren.

Für diese Studie wurden zunächst die Vokabulare und das terminologi-

sche Mapping vorbereitet, die zum Importieren der Faktendaten erforderlich sind. Anschließend wurde ein Prozess zum Importieren der Daten aus den Quelldateien entworfen. Für die acht Universitätskliniken des MIRACUM-Konsortiums wurde eine mit der OMOP-Datenbank und den OHDSI-Tools vorkonfigurierte virtuelle Maschine sowie Werkzeuge zum Datenimport und zur Durchführung der Analyse bereitgestellt. Jeder Partner instanziierte diese virtuelle Maschine lokal und wendete die Importjobs an, um die Daten zu laden. Zuletzt wurde eine beispielhafte föderierte/verteilte Abfrage als Proof-of-Concept für den Ansatz ausgeführt.

Beim Import der Daten bestand die besondere Herausforderung in der Abbildung der Terminologien. Für den Import des MI-I-Kerndatensatzes mussten Zuordnungen in drei Bereichen gefunden werden: ICD-10-GM für Diagnosen, OPS für Prozeduren und individuelle Konzeptcodes z. B. für das Geschlecht des Patienten. Während dies für ICD-10-GM schon gelang, war dies für OPS noch nicht möglich. Als Workaround wurden die OPS-Konzepte als neue Terminologie eingeführt, was für die vernetzte Forschung in Deutschland ausreicht. Für die Teilnahme in internationalen Studien der OHDSI Community muss die Terminologie jedoch noch angepasst werden.

Insgesamt wurden im Rahmen der Studie die Daten von über drei Millionen stationären Patienten mit ca. 26 Millionen Diagnosen und 21 Millionen Prozeduren aus den Jahren 2003 bis 2016 an den acht Standorten importiert. Basierend auf diesen Daten erfolgten bereits erste Auswertungen. So wurde beispielsweise die Behandlungsabfolge (Operation, Bestrahlung, Chemotherapie) bei kolorektalem Karzinom in Abhängigkeit von der Lokalisation des Tumors (Dickdarm, Übergang, Rektum) visualisiert (Abbildung 2). Hierzu wurde an einer Universitätsklinik eine Query auf der Datenbank erstellt, welche an die an-

deren Universitätsklinik verteilt und deren Teilergebnisse zentral zusammengetragen wurden. Im Ergebnis wurden 16 700 Patienten in die Abfrage eingeschlossen. Da nur summarische Ergebnisse und keine individuellen Patientendaten die Standorte verlassen haben, entsprach die Abfrage den geforderten datenschutzrechtlichen Bestimmungen.

#### **ZUSAMMENFASSUNG**

OMOP ist ein umfassendes Datenmodell für Patientendaten sowie eine Sammlung von mehr als 70 Vokabularen. OHDSI ist eine Community, welche OMOP nutzt, weiterentwickelt und eine Reihe von Tools bereitstellt, wie Athena, Atlas, Achilles und verschiedene R-Pakete zur Analyse der Patientendaten von Kohorten.

OMOP wurde in MIRACUM erstmalig in einem deutschen Konsortium von Universitätskliniken auf Basis eines typischen Datensatzes erfolgreich angewendet. Bisherige Erfahrungen zeigen, dass das OMOP CDM und die OHDSI-Werkzeuge gut für die vernetzte Forschung genutzt werden können, weshalb es als Forschungsdatenbank in MIRACOLIX aufgenommen wurde.

Da jedoch die von OHDSI vorgegebenen Terminologien wie SNOMED in Deutschland nicht verwendet werden, ist ein zusätzliches Mapping erforderlich. Dies kann einfach sein, wenn es in Athena bereits ähnliche oder kompatible Terminologien gibt, könnte jedoch auch zu großen Aufwänden führen, wenn dies nicht der Fall ist. ■



■ **PROF. DR. MARTIN SEDLMAYR**

Institut für Medizinische Informatik und Biometrie, Medizinische Fakultät Carl Gustav Carus, TU Dresden  
Kontakt: martin.sedlmayr@tu-dresden.de

■ **PROF. DR. HANS-ULRICH PROKOSCH**

Lehrstuhl für Medizinische Informatik, Friedrich-Alexander-Universität Erlangen-Nürnberg