

OneMII Symposium Chemnitz

Mit dem Datenhotel zur sicheren Nutzung klinischer Routinedaten

Dr. Layla Riemann, 01.10.25
SMITH-Nachwuchsgruppe BENEFIT & AG KIP

Die Arbeitsgruppen

BENEFIT

- Nachwuchsgruppe am UKE, gefördert durch das BMBF/BMFTR
- Ziel: Brücke zwischen klinischer Versorgung und Forschung
- Fokus: Datenharmonisierung, Standardisierung, Anforderungen aus der Forschung



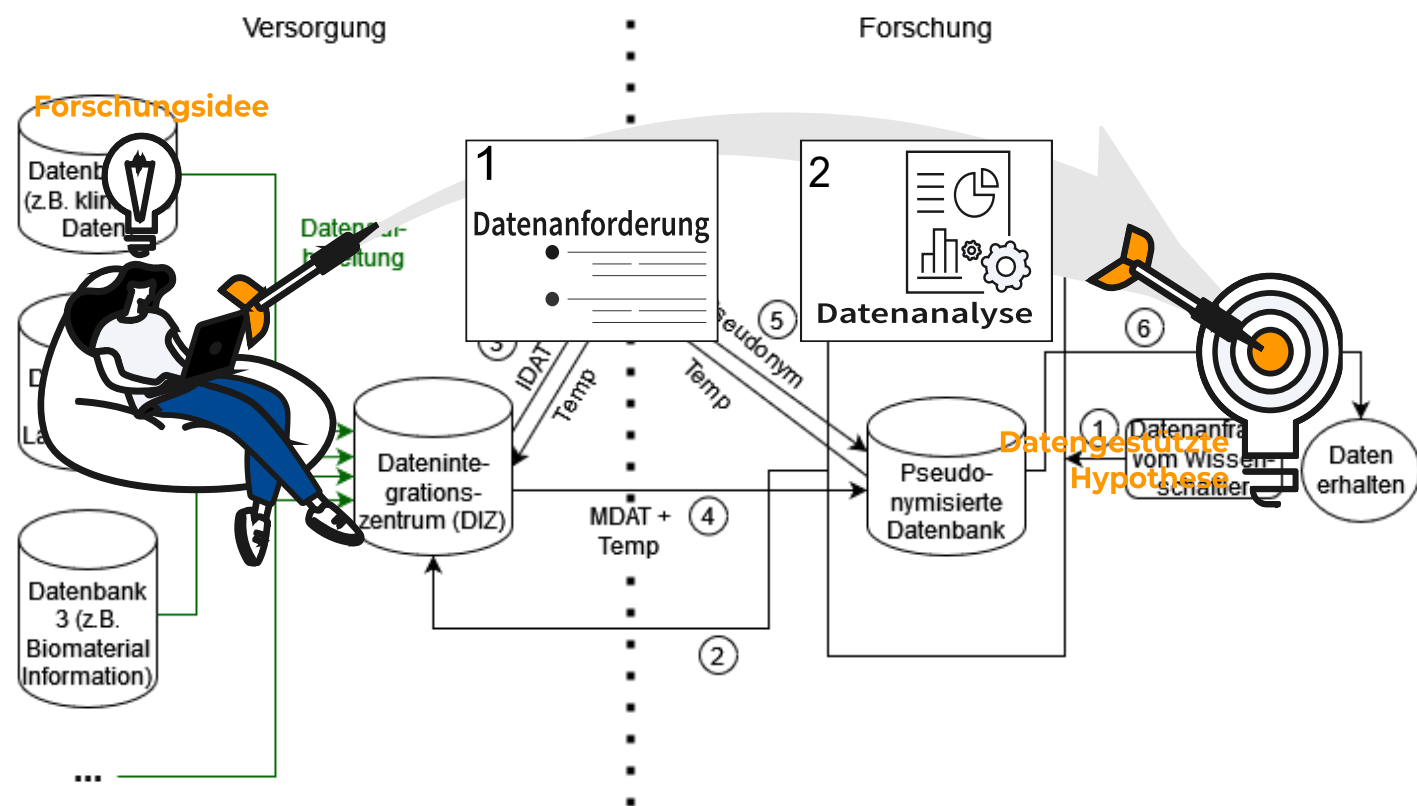
AG KIP

- Institut für angewandte Medizininformatik, UKE
- Projektziel: Nutzerfreundliche Datenauswahl und Analyse, multimodale Integration von Datenquellen
- Fokus: KI-gestützte Visualisierungen, synthetische Daten, LLM-Integration in der Onkologie



Das Datenhotel – Konzept & Status Quo

- Kernidee: Klinische Routinedaten (ab 2008-heute) retrospektiv und sicher nutzbar machen
- Rechtliche Grundlage: § 12 Hamburgisches Krankenhausgesetz
- Technisches Prinzip:
 - Automatisierter Export aus KIS & PDMS
 - Pseudonymisierung durch Treuhandstelle
 - Bereitstellung in einer Jupyter Notebook Umgebung (Python, R, Julia)
- Funktionen für Forschende:
 - Browserbasierte Datenselektion
 - Interaktive Analysen und Visualisierungen
 - Pilot-Fallstudien & Evaluationen



Abarbeitung einer wissenschaftlichen Anfrage

- Fragestellung der Studie:



„Untersuchung des Einflusses hämodynamischer Parameter und Beatmungsdrücke auf das postoperative Outcome bei herzchirurgischen Patienten.“

- Datenquellen:
 - KIS (Krankenhausinformationssystem)
 - PDMS Intensivstation
 - PDMS OP
 - Geräte-Datenströme
 - Alles im Datenhotel vorhanden



Stolpersteine im Detail

1. Unterschiedliche Zeitaufösungen



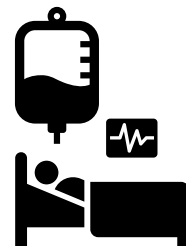
- Blutdruck im KIS: 3x täglich dokumentiert
 - Blutdruck im PDMS Intensiv: alle 15 Minuten
 - Blutdruck im OP-PDMS: alle 5 Minuten
 - Gerätedaten (z. B. invasive Arterienmessung): bis zu 500 Hz
- Problem: Harmonisierung unmöglich ohne massive Aggregation oder Verlust von Information.
 - Frage: Welches „Granularitätsniveau“ ist für die Forschung korrekt?



Stolpersteine im Detail

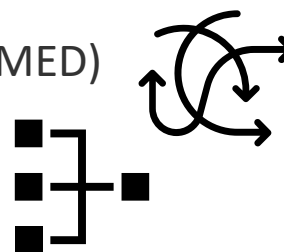
2. Unterschiedliche Geräte und Namenskonventionen

- Laborparameter wie Kreatinin oder Hämoglobin aus verschiedenen Analysegeräten → leicht abweichende Normwerte, Messmethoden
- Gleich benannte Variablen haben unterschiedliche Bedeutung (z. B. „Lactat“ aus BGA vs. Zentrallabor)



- Problem: Mappings erforderlich (Ontologien, LOINC/SNOMED)

- Gefahr: Scheinbar gleiche Werte sind nicht vergleichbar.



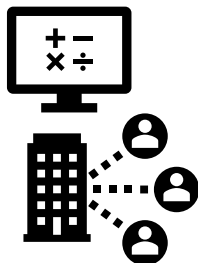
- Lösung: Interdisziplinäre Annotationsteams (Ärzte + Informatiker + Data Scientists) und Data Stewards / Bereichs-Datenexperten

- Rewarding System / Refinanzierung dieser Personalressourcen
- Schulung der Forschenden in medizinischem Datenverständnis
- Erzeugen von abfragbaren und wieder nutzbaren „Phänotypen“
- Aufbau von Wissensgraphen zur Kontextualisierung



Stolpersteine im Detail

3. Kontextabhängigkeit klinischer Variablen



- Beatmungsdruck: Gerät A dokumentiert *P_{insp}*, Gerät B nur *Driving Pressure*
- Interpretation, ob *Driving Pressure* = *P_{insp}* – *PEEP* oder separat dokumentiert → geräteabhängig
- Ohne Kenntnis aller Variablen (*PEEP*, *P_{insp}*, Gerätetyp) nicht interpretierbar

- Problem: Daten ohne klinisches Wissen führen zu falschen Schlussfolgerungen.



- Wird größer, je mehr Quellen es gibt (PACS, ICM, Medikamente etc.)

- Lösung: Aufwändiges Preprocessing bzw. abteilungsabhängige Hilfe von Data Stewards nötig



Stolpersteine im Detail

4. Dokumentationspraktiken & menschlicher Faktor



- KIS-Einträge oft von Pflegekräften/Ärzten stichpunktartig dokumentiert → subjektiv und zeitversetzt
- Routinedaten ≠ Forschungsdaten
- Dokumentation primär für Versorgung, nicht für Wissenschaft
- Heterogene Qualität je nach Fachsystem und Anwender:innen
- PDMS zieht Gerätedaten automatisch → objektiv, aber „roh“ und teilweise unvalidiert
- Fehlende Werte werden im KIS manchmal geschätzt oder kopiert

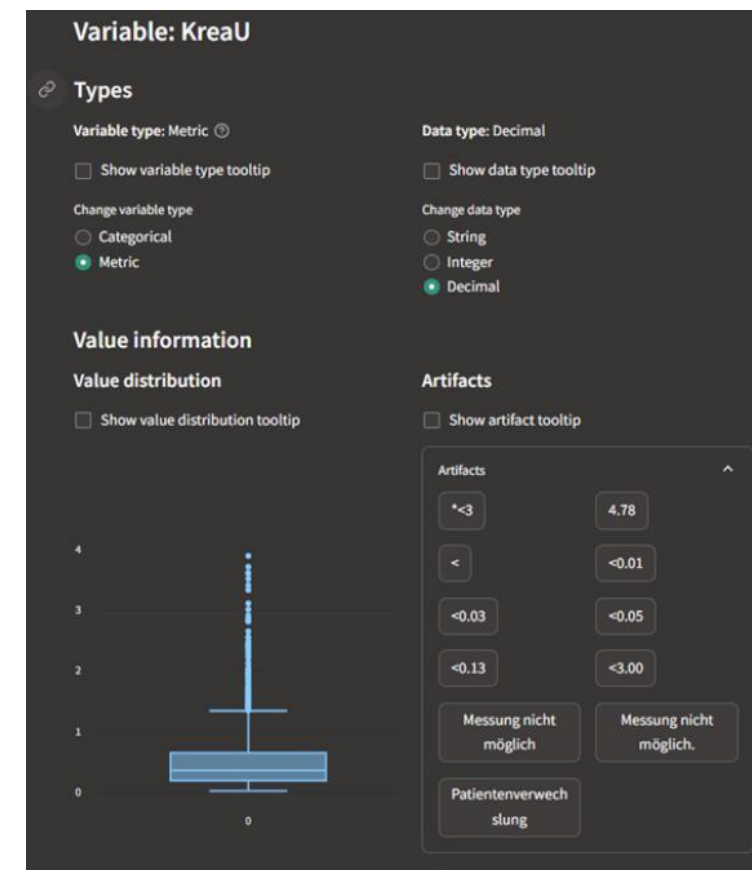
- Problem: Vermischung von „gelebter Dokumentation“ und „harter Messung“

- Offene Fragen:

- Wie wird Datenqualität für Forschende sichtbar gemacht?
- Welche Validierungsmechanismen sind etabliert?

- Lösungen:

- Systematische Qualitätsmetriken (Vollständigkeit, Konsistenz, Plausibilität)
- Transparente Metadatenkataloge
- Feedback-Schleifen mit klinischen Anwender:innen



Stolpersteine im Detail

5. Unterschiedliche Speicherorte & Versionen



- OP-PDMS und Intensiv-PDMS speichern dieselben Beatmungsdaten → aber in unterschiedlichen Datenbanken, mit leicht abweichender Struktur
- Laborwerte: Zwischenspeicherung im LIS, später ins KIS exportiert → Zeitstempel-Differenzen

- Problem: Welche Version ist die „wahre“?



- Lösung: Aufwändiges Preprocessing bzw. abteilungsabhängige Hilfe von Data Stewards nötig



Stolpersteine im Detail

6. Neue Wissenschaftskultur – Begehrlichkeiten & Zugang



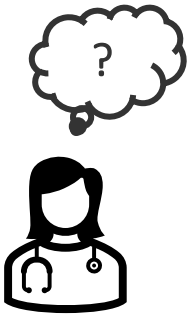
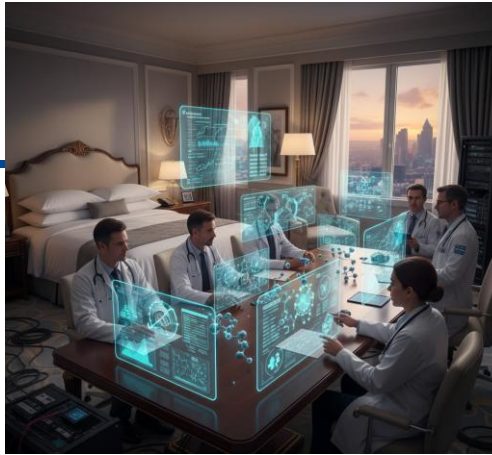
– Wenn viele Forschende auf dieselben Daten zugreifen, entsteht Konkurrenz um Themen („first to publish“)

- Risiko: Fragmentierung von Forschung & Mehrfachauswertungen ohne Abstimmung

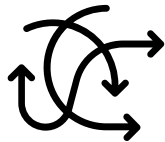
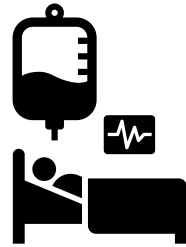
- Lösungen:

- Einrichtung von Koordinationsgremien für Datenzugang
- Dokumentation von Anfragen & Analysen (Register)
- Förderung von Kollaboration statt Konkurrenz

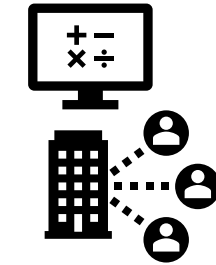




Zeitskalen



Namens-
konventionen



Kontext



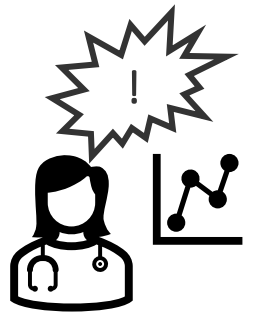
Dokumentation



Speicherorte



Publikation



Weitere offene Herausforderungen

- Echtzeitanalysen & Datenpipelines
 - Aktueller Stand
 - Unterschiedliche Daten-Zugriffsmöglichkeiten je nach Fachsystem
 - Teilweise tägliche Aktualisierung (24h) von Datenbankkopien → Reporting & Dashboards
 - Limitierungen
 - Nicht geeignet für klinische Entscheidungsunterstützung in Echtzeit
 - Verzögerungen → Forschung mit aktuellem Datenstand oft nicht möglich
 - Wünschenswert
 - Echtzeit-Pipelines (z.B. Apache Kafka, Streaming-Architekturen)
 - Near-Real-Time Dashboards für Monitoring und Forschung
 - Abgrenzung: Forschung vs. unmittelbare Patientenversorgung

Weitere offene Herausforderungen

- Trusted Research Environments (TRE) & Secure Processing Environments (SPE)
 - Aktueller Status
 - TRE/SPE am UKE → retrospektive Datennutzung im Datenhotel
 - Starke Sicherheitsarchitektur, aber eingeschränkte Flexibilität
 - Zukünftige Anforderungen
 - Integration von Cloud-Architekturen (Skalierung, Rechenpower)
 - Unterstützung für föderiertes Lernen und Multi-Site-Analysen
 - Erweiterte Tools: GPU-Cluster, S3-Storage, Containerisierung (Docker/Kubernetes)
 - Chancen
 - Erschließung komplexer Analysen (KI, Bilddaten, Omics)
 - Harmonisierung mit europäischen SPE-Initiativen (z. B. FDPG, EHDS)

AG „Knowledge Integration in Precision Medicine“



Vielen Dank!



**BENEFIT: Nutzenoptimierung und
Verfügbarkeitserhöhung von
Gesundheitsdaten**



Dr. Layla Riemann

Gruppenleitung „Knowledge Integration in Precision Medicine“, Institut für Angewandte Medizininformatik (IAM)

Christoph-Probst-Weg 1 | D-20251 Hamburg

Telefon +49 (0) 40 7410-57491

l.riemann@uke.de | www.uke.de

Dr. Philipp Breitfeld

Gruppenleitung BENEFIT, IAM, Zentrum für Anästhesiologie und Intensivmedizin, Klinik und Poliklinik für Anästhesiologie

Martinistraße 52, Gebäude Ost 10 | 20246 Hamburg

Telefon +49 (0) 40 7410-39023

p.breitfeld@uke.de | www.uke.de