

# Der methodische Use Case PHEP

## Von Daten über Informationen zu Wissen - die Nahtstelle zur klinischen Forschung

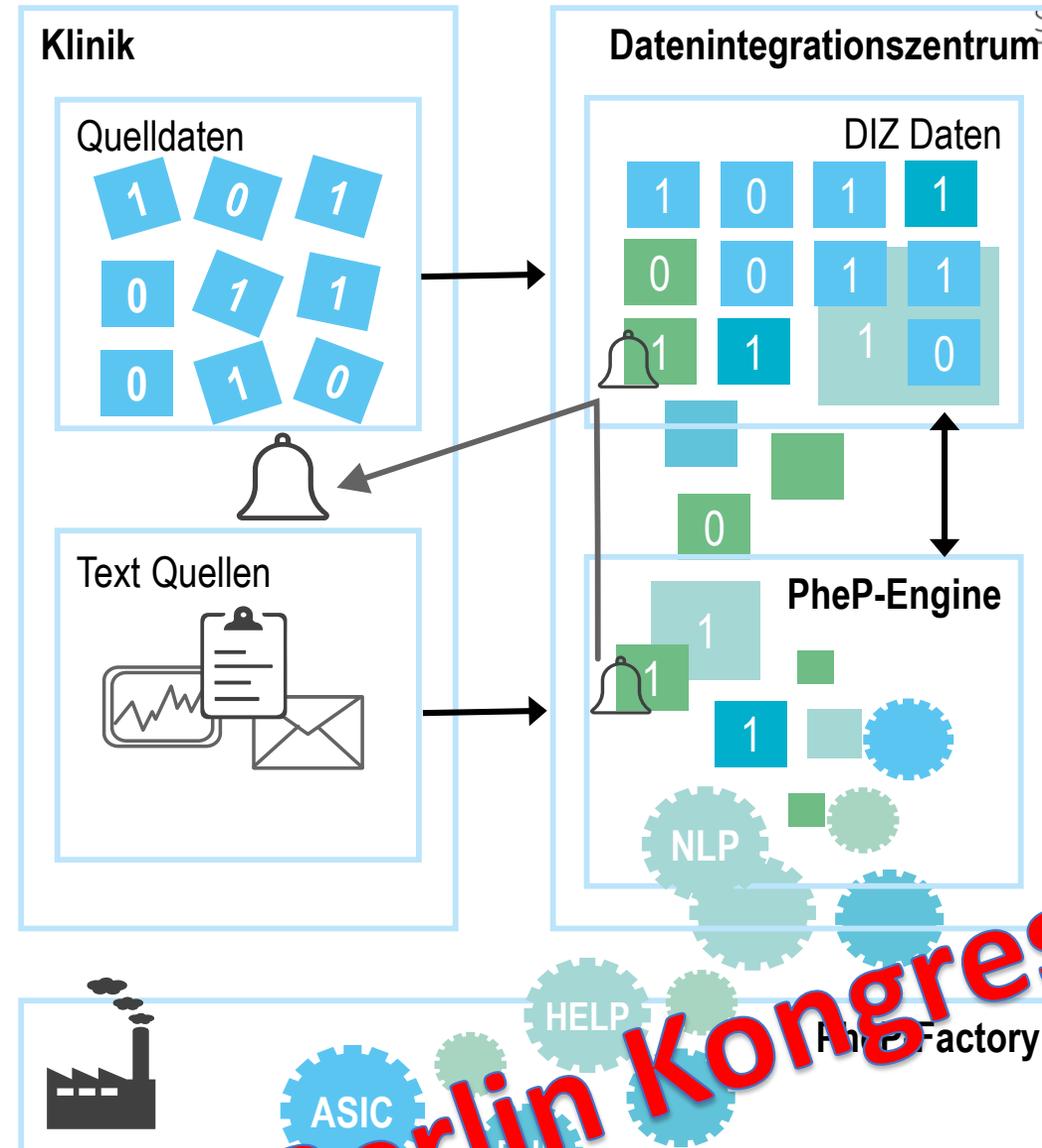


Dr. Frank A. Meineke  
Institut für medizinische Informatik, Statistik und Epidemiologie  
Universität Leipzig

*New Solutions in Digital Health*  
*dbb forum Berlin 2022*

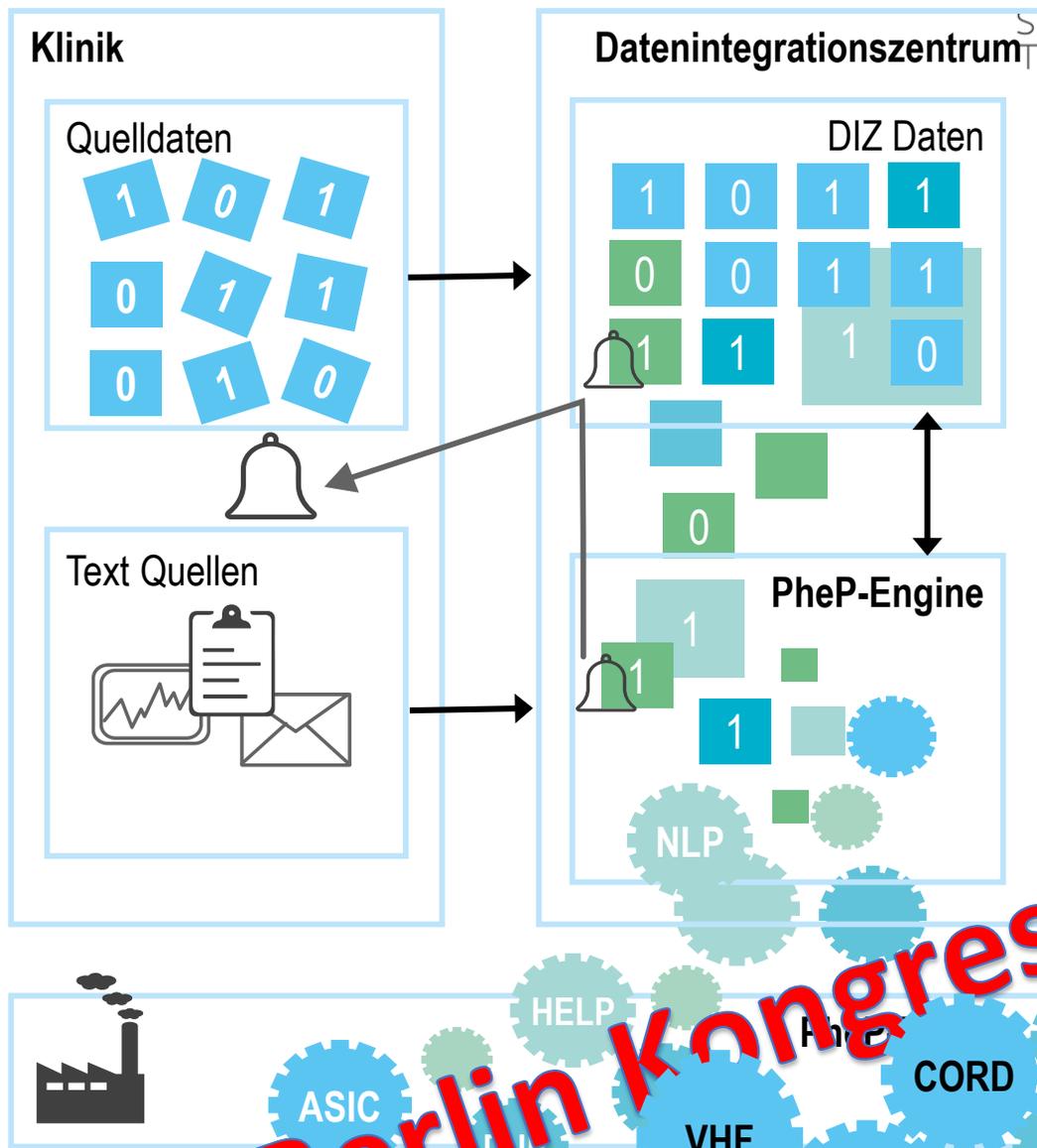
# Einbettung des Phenotyping

- Daten Transformation  
ETL, FHIR
- PheP-Engine  
Data processing algorithms
- Phänotypberechnung  
auch: Kuratation, Kohorten/Gruppen
- NLP Text Quellen  
Text → strukturierte Daten
- Alarme/Rückmeldungen  
erzeugt durch spez. Phänotypen
- PheP-Factory  
*Ideen → Algorithmen*



# Einbettung des Phenotyping

- Daten Transformation  
ETL, FHIR ✓
- **PheP-Engine**  
Data processing algorithms
- **Phänotypberechnung**  
auch: Kuration, Kohorten/Gruppen
- **NLP Text Quellen**  
Text → strukturierte Daten
- Alarme/Rückmeldungen  
erzeugt durch spez. Phänotypen
- PheP-Factory / **DUPS**  
*Ideen → Algorithmen*



**Berlin Kongress 2019**

# Use Projects innerhalb der MII

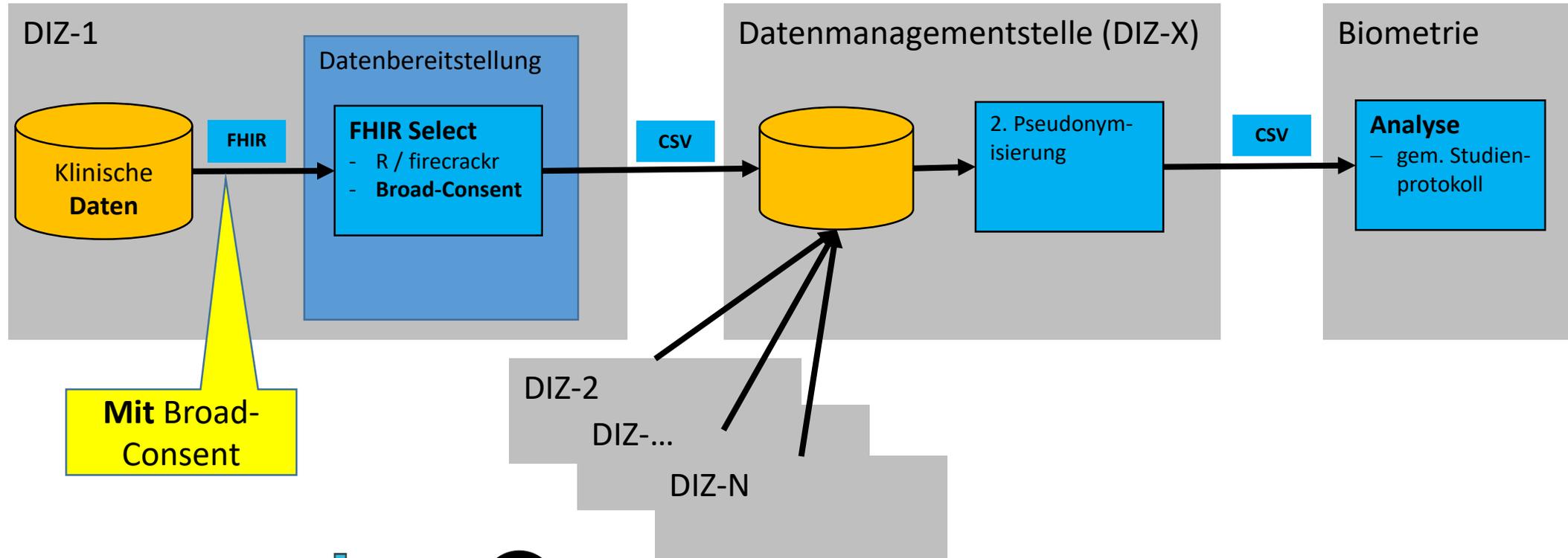
- Laufend: Standort bezogene Datennutzungsprojekte
  - Heterogene Anwendungen und Ausleitungsverfahren
- 2018: Demonstratorstudie
  - Zusammen mit Projektstart
  - Machbarkeit der Zusammenarbeit, Solitär / kein Regelbetrieb
- Ab 2018: Use Cases der einzelnen Konsortien (SMITH: **ASIC, HELP**)
  - Zeitgleich mit DIZ-Aufbau
  - hoher Grad an Abweichung vom geplanten Regelbetrieb
- Ab 2020: Konsortial übergreifende Use Cases (**POLAR, CORD, ABIDE, CODEX**)
  - Frühe Phase im DIZ-Aufbau
  - Flexibilität und Kompromissbereitschaft auf Nutzerseite
- 2021-22: MII weiter 6. Projektathon (**WE-STORM, VHF\_MI**)
  - Endphase DIZ-Aufbau
  - Keine Kompromisse; Regelbetrieb wird erwartet

# Übergreifende Use Cases testen das Gesamtsystem der MII

- **Zentrale Vorgaben der MII** und Dissemination in die DIZ
  - Kerndatensatz, Data Sharing Framework, Forschungsdatenportal Gesundheit, Broad Consent
- **Organisationsabläufe der Kliniken** (inkl. Apotheken und Labore)
  - Was wird wann dokumentiert?
- Ausstattung und **Digitalisierungsgrad der Kliniken** (Existenz der Daten)
  - Wie und mit was wird Dokumentiert?
- **Kooperation Kliniken mit den DIZ** (Daten → DIZ) inkl. klinischer Expertise
  - Wie eng arbeiten Kliniker / Klinik-IT und DIZ zusammen?
- **ETL Strecken** der DIZ (Daten → FHIR)
- **Ausstattung/IT-Infrastruktur/Schulung** der DIZ
- **Bereitstellung** der Daten (FHIR → Datennutzungsprojekt)
- **Qualität der Daten** (nach/bei Auswertung)

# PheP-Engine Pipeline für „Zentrale Analyse“

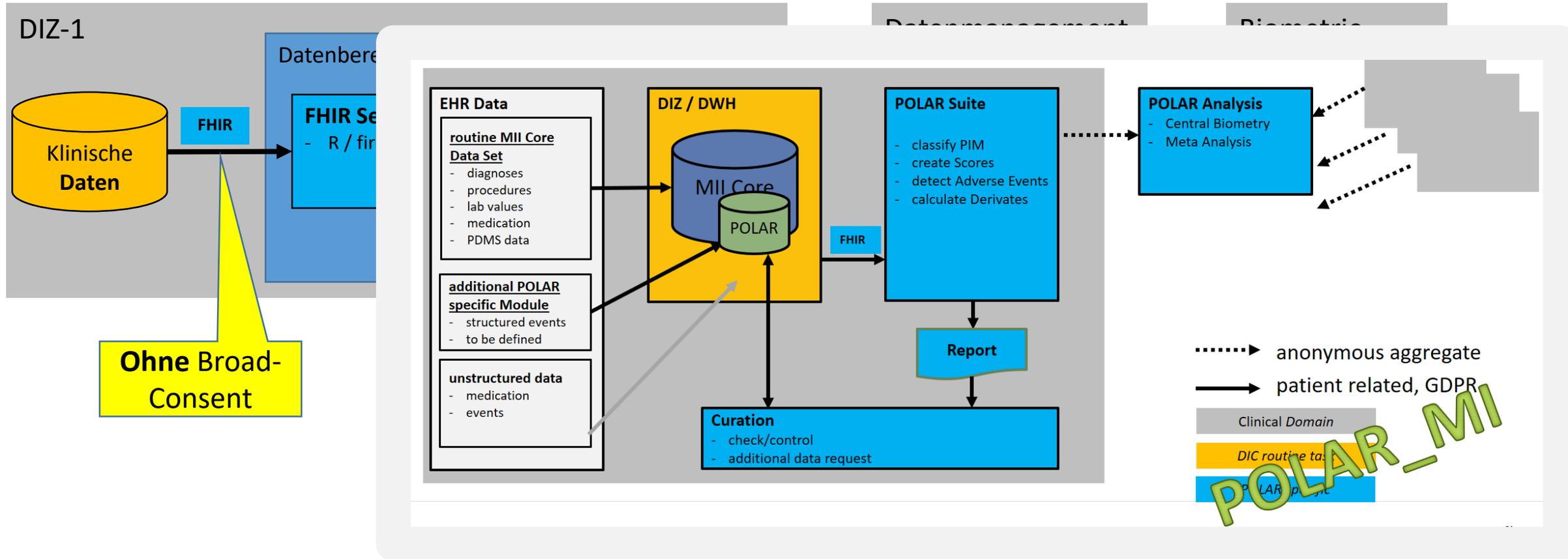
Erster Durchlauf im 6. Projektathon als  
SMITH-2 VHF\_MI



—————> Einzeldaten/pseudonym

# PheP-Engine Pipeline für „Dezentrale Analyse“

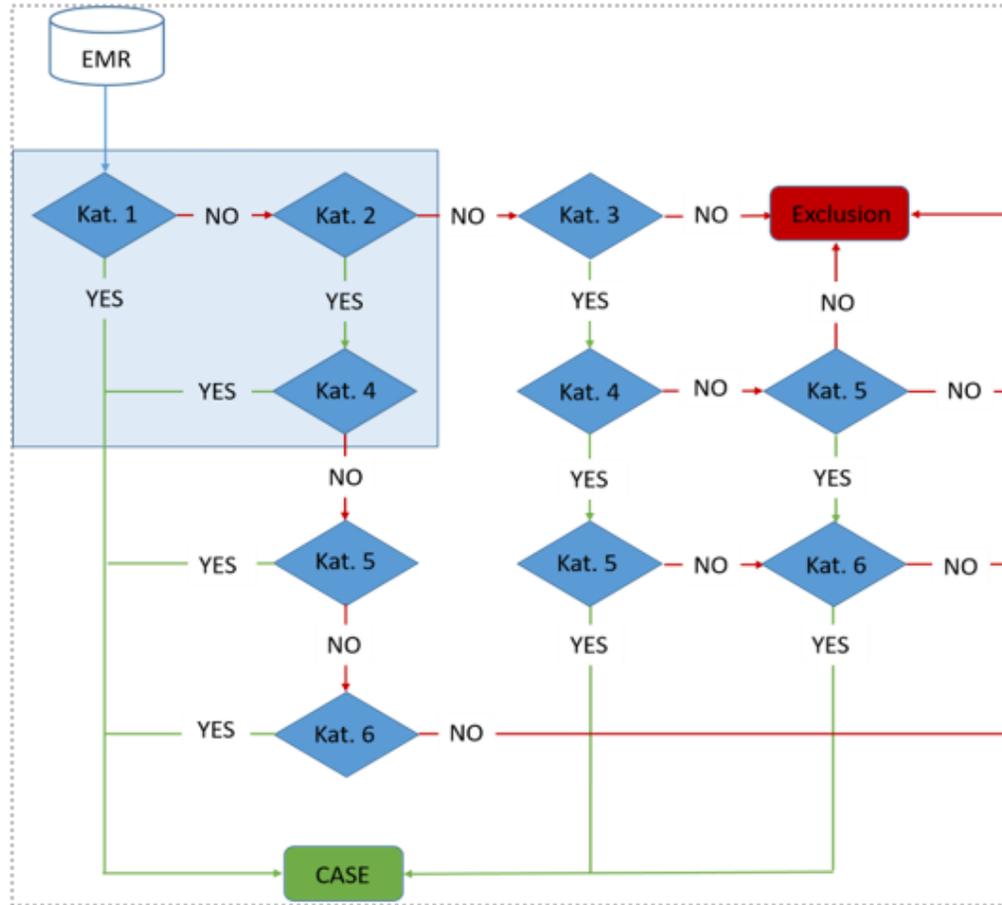
Prototypisch u.a. in POLAR\_MI umgesetzt



# Phänotypisierung Fortführung 2021ff

Term

## Flow Chart Sturz



### Mögliche Merkmalkategorien:

**Kat. 1:** ‚starke‘ ICD-10-Codes  
(wichtigste/häufigste Frakturen: Hüfte,  
Oberschenkel, Handgelenk, Schädel)

**Kat. 2:** ‚schwächere‘ ICD-10-Codes  
(oberflächliche Verletzungen, offene  
Wunden, sonstige Frakturen)

**Kat. 3:** unspezifische ICD-10-Codes  
(sonstige Verletzungen; Verletzung nicht  
näher bezeichnet; sonstiger Unfall)

**Kat. 4:** OPS-Codes (z. B. CT, MRT,  
Reposition von Fraktur und Luxation oder  
Z01.6 (Röntgenuntersuchung, ICD-10))

**Kat. 5:** Grunderkrankungen (z. B. Demenz,  
M. Parkinson, Sehstörungen, Schwindel)

**Kat. 6:** Alter ≥ 65 Jahre

erung



Jciteli

ung

quellen

in  
bar und

gorithmen  
sitory

Algorithmen

Terminology- and Ontology-based Phenotyping (TOP) Framework

# Phänotypisierung Fortführung 2021ff

## KI-basierte Phänotypisierung

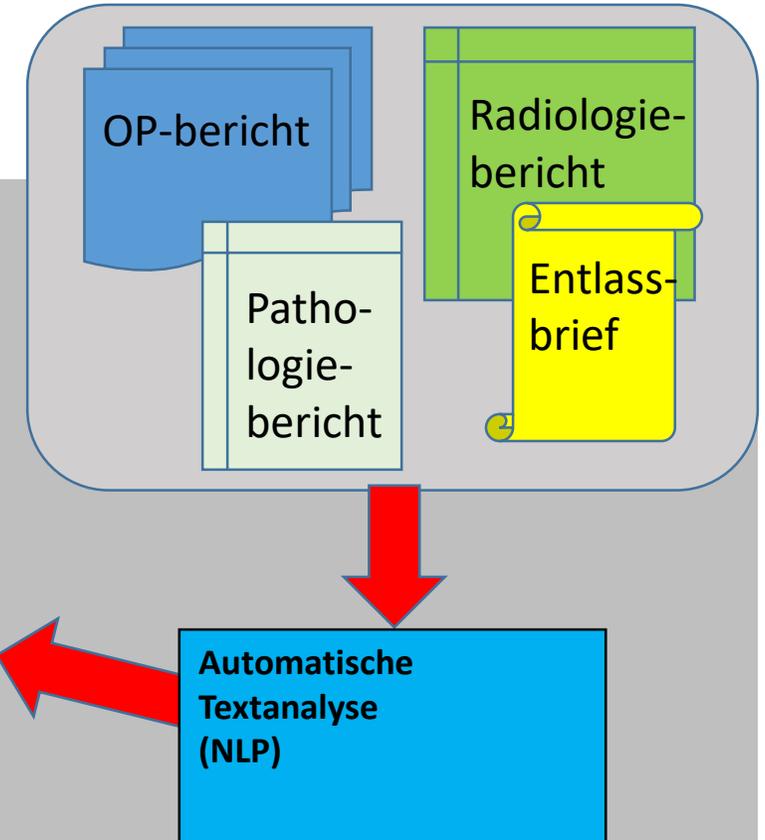
- Kohorten = Patientengruppe, die sich anhand spezifischer Eigenschaften ergibt
  - z.B. Alter > 50, (Haupt-/Neben-) Diagnose = Diabetes Typ 2 etc.
- Nutzung von Methoden des Data Mining und des Maschinellen Lernens zur datengetriebenen Profilierung
  - Aufdecken von Klassen (Clustering)
  - Klassifikation von Patienten, z.B. entsprechend Ihrer Symptomatik
  - Prädiktion von zukünftigen Ereignissen
- Nutzung spezieller Methoden des verteilten Rechnens
  - Personal Health „Algorithm to the Data“
- Professur für Medical Data Science  
**Toralf Kirsten**
- DUPs zur Nutzung von Versorgungsdaten über MII Infrastruktur
  - Einfluss der standardisierten Dokumentation auf Behandlungsentscheidungen in der Ophthalmologie
  - Unterscheidung von Diabetes Typ 2 Patienten
  - Risikofaktoren von Patienten vor und während chirurgischer Eingriffe
  - ...

# PheP/NLP

## Strukturierung von Daten

Klinik → DIZ

Pat-ID	Name	Alter	Geschlecht	Diagnose (ICD)	Medikation	...
UH-5402	Pfeiffer	68	W	G20.11	Candesartan	
SK-3390	Auer	34	M	K45.8	Prednison	
PJ-1095	Strosz	91	M	I20.1	Amlodipin	
ST-4793	Zencik	23	W	070.1	Fentanyl	
.....						



# PheP/NLP

## Leistungsspektrum NLP in SMITH

- **Anonymisierung** und Pseudonymisierung:
  - Erkennung und Änderung individuen-identifizierender Merkmale (Name, Adresse usw.)
- **Sektions**
  - Familien
- **Erkennung**
  - Medik
- **Erkennung**
  - Medik
- **Erkennung**
  - X-Test
- **Erkennung der Sicherheitsgrade & Negation zu klinischen Aussagen.**
  - X-Test indiziert sicher Krankheit, Patient diagnostiziert-mitvermutlich Krankheit usw.

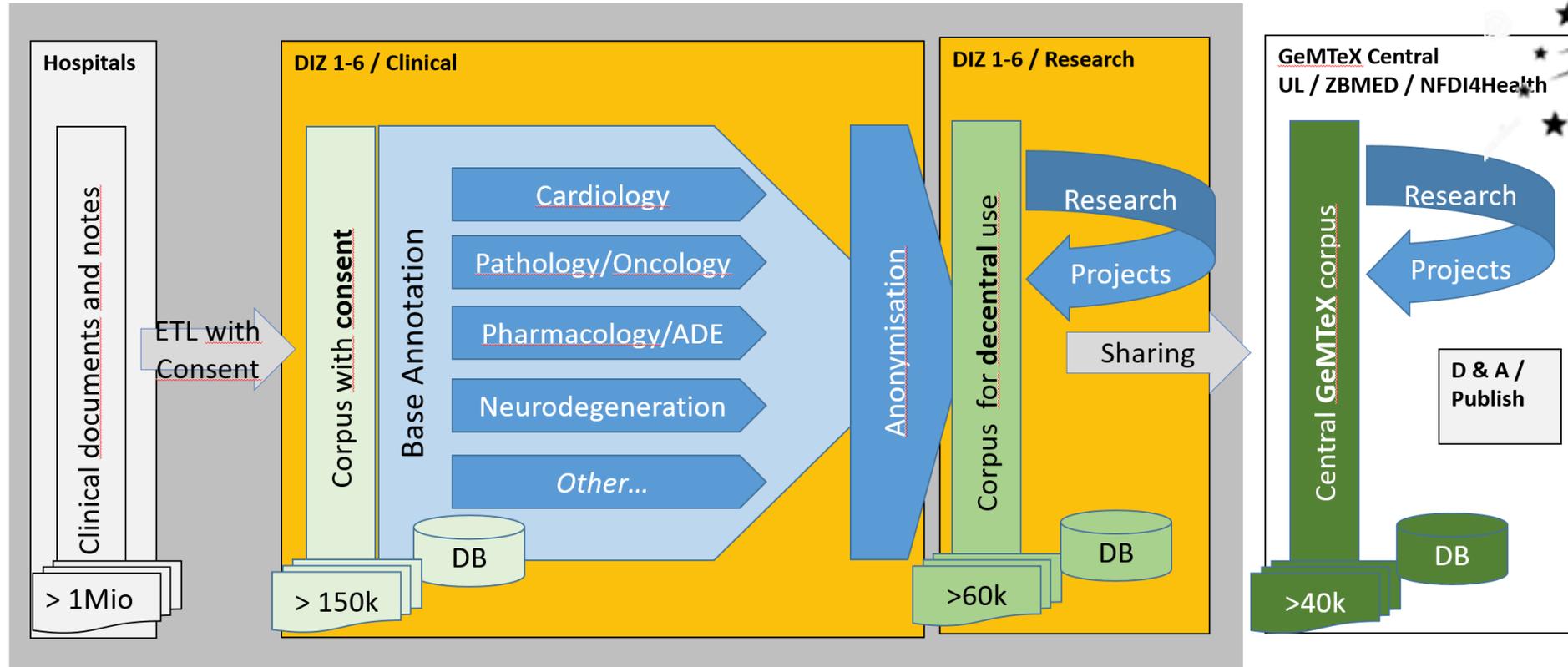
Flaschenhals: Forschung auf **deutschsprachigen** Korpus  
 Basis hier: 1000 Patientenakten Projekte SMITH

Noch fehlend: Umsetzung in Konzepte der MII  
 - FHIR Präsentation, Dokumente, Provenienz, Ergebnisse/Sicherheiten

→ „Dickes Brett“

# PheP/NLP Fortführung 2023ff

## GeMTeX – German Medical Textcorpus



# Lessons Learned

- Kliniken und DIZ sind und bleiben hinter ihrer interoperablen Äußerem intern jeweils einzigartig und individuell
  - Achtung: DIZ können nur Daten bereitstellen
- Use Cases treiben die Entwicklung!
  - Z.B. POLAR → Medikation; aber: DIZ erzeugen keine Daten -- sie stellen sie bereit
- Die übergreifende Datennutzung in der jetzigen Anlaufphase deckt verbliebene Inkompatibilitäten auf
  - Problembereiche: KDS Umsetzung, Infrastruktur/FHIR, Software, Semantik, Organisation
- Broad Consent noch nicht in der Breite gelebt
  - Projektathon SMITH-2 = VHF Patientenzahlen noch viel zu klein
- „Probleme sind dornige Chancen“ 😊
  - Z.B. Auswertung Projectathon, DUP VHF-SMITH „7 von 7“

Aber: die erzielte Vernetzung und offene interdisziplinäre Zusammenarbeit durch die MII ist rückblickend **revolutionär**.

# Zusammenfassung und Ausblick

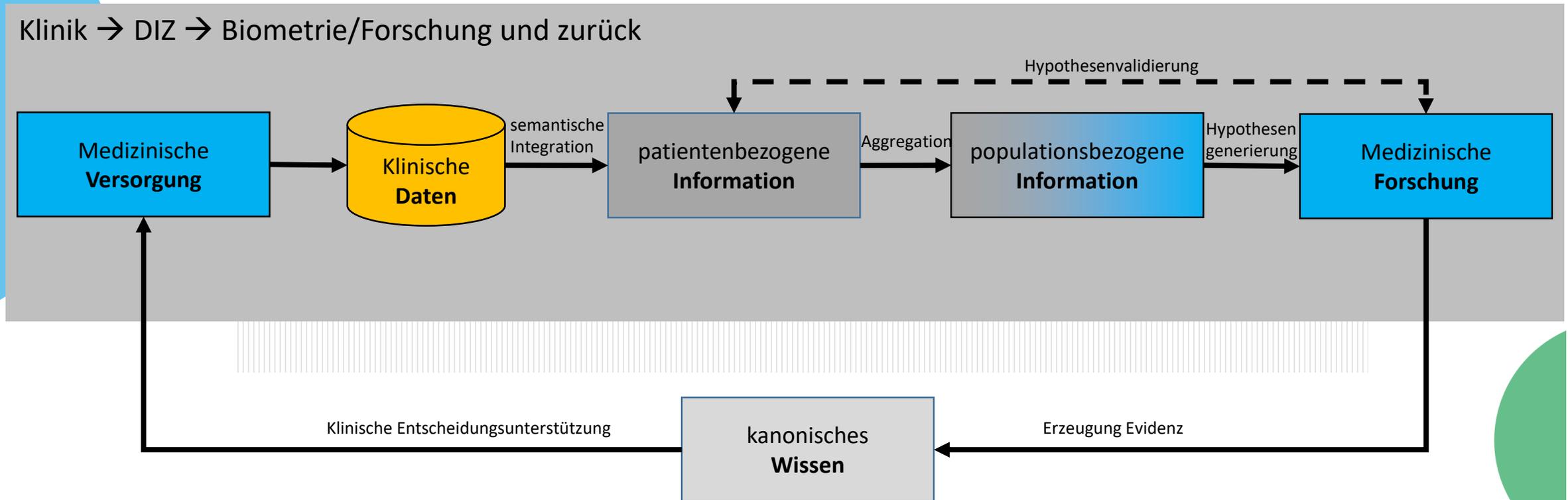
## Methodischer Use Case PheP 2023ff

- Use Cases (*DUP*), nicht Datensammlung, stehen nun im Mittelpunkt
- Grenzen zwischen den Konsortien werden durchlässig
  - Konsortiale DUPs ein Auslaufmodell
- Der Methodische Use Case PheP differenziert und integriert sich
  - Optimierung der Ausleitungs-/ Analysepipeline („PheP-Engine“) → MII Taskforce DDT, AG DataSharing, Zentrale / Dezentrale Analysen, Methodik (Ad hoc, PHT, DataSHIELD,...)
  - Phänotypisierung/Ontologie → Nachwuchsforschungsgruppe TOP, Professur Medical Informatics
  - Natural Language Processing → GeMTeX
  - Ausbildung → SMITH Akademie
  - Publikationsplattform „SMITH Health Atlas“ → NFDI4Health
- Intensivierung / Fortführung auch in Modul 2 Anträgen der MII
  - Z.B. Medical Device Regulation konforme Umsetzung von Interventionen

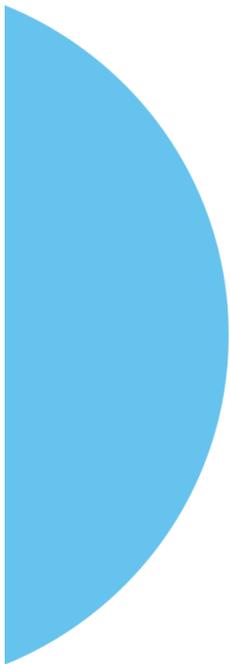


# Der methodische Use Case PHEP

## Von Daten über Informationen zu Wissen - die Nahtstelle zur klinischen Forschung



*New Solutions in Digital Health*  
*dbb forum Berlin 2022*



# Vielen Dank für Ihre Aufmerksamkeit!

**SMITH.**  
**Klinische Forschung und Patientenversorgung nachhaltig verbessern.**

[www.smith.care](http://www.smith.care)

