

Projekt Krebsforschungsdatenzentrum | onkoFDZ

KI-gestützte Evidenzgenerierung aus
versorgungsnahen Daten Klinischer Krebsregister,
GKV-Routinedaten, Klinikdaten und deren Linkage

Prof. Dr. Monika Klinkhammer-Schalke

Gefördert durch:



Bundesministerium
für Gesundheit

aufgrund eines Beschlusses
des Deutschen Bundestages

Förderkennzeichen:
ZMI5-2522DAT14A-O

- Projektstruktur
- Projektübersicht
- Projektidee
- Projektbeteiligte und Kooperationspartner
- Fragestellung
- Projektziele
- KI im Projekt
- Datengrundlage
- Methodische Umsetzung Linkage
- Was haben wir erreicht?
- Wie geht es weiter?

Projektstruktur

Projektleitung und Verbundkoordination:



Prof. Dr. med. Jochen Schmitt, MPH

Direktor des Zentrums für Evidenzbasierte Gesundheitsversorgung (ZEGV),
Universitätsklinikum und Medizinische Fakultät
Carl Gustav Carus an der TU Dresden
Tel.: 0351 / 458 64 95
E-Mail: Jochen.Schmitt@ukdd.de



Prof. Dr. med. Monika Klinkhammer-Schalke

Past President
Arbeitsgemeinschaft Deutscher Tumorzentren e.V. (ADT)
Direktorin Tumorzentrum Regensburg, Zentrum für
Qualitätssicherung und Versorgungsforschung an
der Fakultät für Medizin der Universität Regensburg
Tel.: 0941 / 943 18 03
E-Mail: Monika.Klinkhammer-Schalke@ur.de

Ansprechpersonen Projektkoordination:



PD Dr. rer. nat. Olaf Schoffer

Bereichsleiter onkologische Versorgungsforschung
am ZEGV
Tel.: 0351 / 458 64 94
E-Mail: Olaf.Schoffer@ukdd.de



Bianca Franke

Geschäftsstellenleiterin
Arbeitsgemeinschaft Deutscher
Tumorzentren e.V. (ADT)
Tel.: 030 / 326 787 27
E-Mail: Franke@adt-netzwerk.de

Projektidee

KI-gestützte Evidenzgenerierung aus versorgungsnahen Daten Klinischer Krebsregister, GKV-Routinedaten, Klinikdaten und deren Linkage

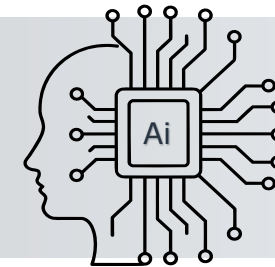
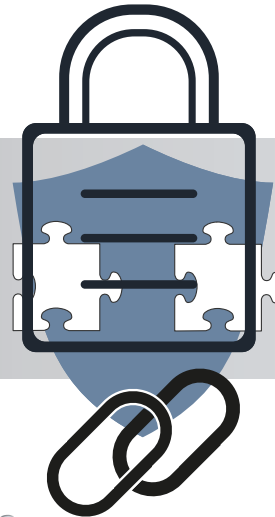
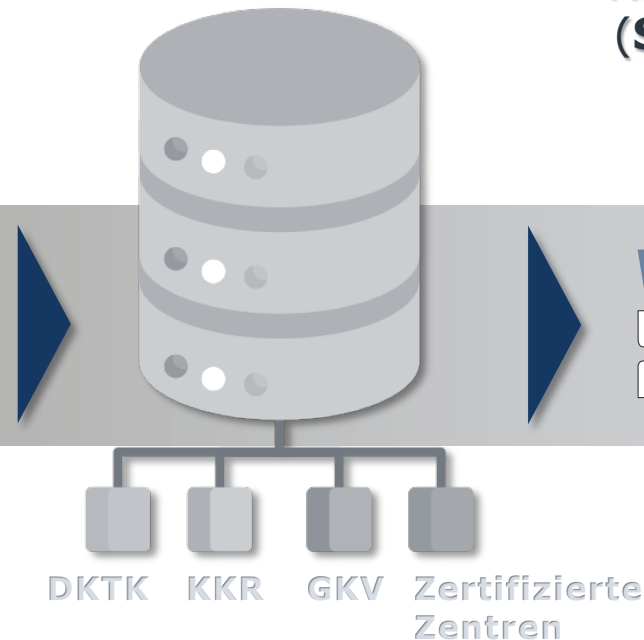
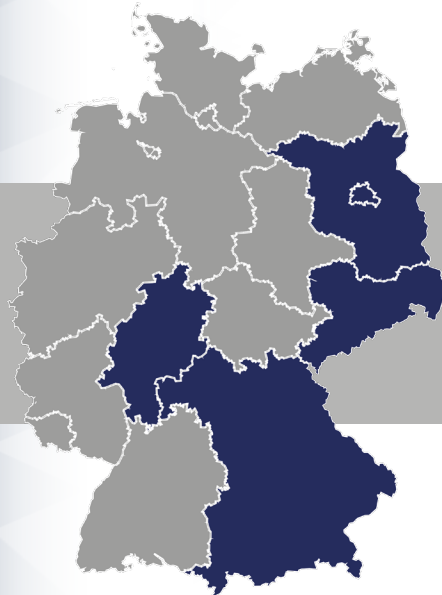
4 Bundesländer
15 Partner

4 verschiedene
Datenquellen

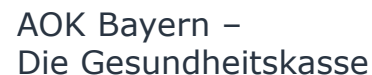
Proof of Concept
Secure Multi-Party
Computation
Privacy-Preserving
Record Linkage
(**SMPC-PPRL**)

Auswertung
mit
KI

Evidenzgenerierung
aus Realdaten



Projektbeteiligte und Kooperationspartner



Arbeitsgruppenleitung

AG Datenbereitstellung

- Herr Kees Kleihues-van Tol (ADT)
- Herr Prof. Dr. Martin Lablans (DKFZ)
- Herr Dr. med. Michael Gerken (TZR)

AG Klinische Fragen und Evidenz

- Herr Dr. med. Vinzenz Völkel (TZR)
- Herr Prof. Dr. Stefan Benz (addz e.V.)
- Frau Dr. med. Judith Hansinger (TZR)

AG KI

- Frau Prof. Dr. Dr. Melanie Börries (Uni Freiburg)
- Herr Prof. Dr. med. Jörg Janne Vehreschild (Uni Frankfurt)

AG Datenauswertung

- Herr PD Dr. Olaf Schoffer (ZEGV)
- Frau Dr. med. Soo-Zin Kim-Wanner (Hessisches Krebsregister)
- Herr Dr. Tobias Kussel (DKFZ)

AG Koordination

- Frau Bianca Franke
- Herr PD Dr. Olaf Schoffer (ZEGV)

- Einige medizinische Fragestellungen können nicht oder nur schwer durch randomisierte kontrollierte Studien (RCTs) beantwortet werden.
 - Dies führt zu Evidenzlücken, insbesondere bei:
 - Speziellen Patientengruppen (z.B. Hochbetagte, multimorbide Patienten)
 - Komplexen Interventionen, die nicht leicht in kontrollierten Studien getestet werden können
- Erhöhung der Unsicherheit bei der Entwicklung von Leitlinienempfehlungen
- Beispiel „**Kolorektales Karzinom**“

Fragestellung

Datenauswertung für Use Case Kolorektales Karzinom

- Analyse folgender Beispielfragestellungen (ohne Leitlinienempfehlung)
 - a) Adjuvante Chemotherapie Kolonkarzinom im Alter > 75 Jahre*
 - b) Adjuvante Chemotherapie Kolonkarzinom in UICC Stadium II*
 - c) Adjuvante Chemotherapie Rektumkarzinom nach neoadjuvanter Vorbehandlung*
 - d) Wahl der systemischen Therapie in Abhängigkeit von der molekularpathologischen Subgruppe und der Tumorlokalisation*
 - e) Laparoskopische Operation und robotische Chirurgie beim Kolon- und Rektumkarzinom*

Es ist anzunehmen, dass sich bei fehlender Leitlinienempfehlung Unterschiede im Behandlungsverhalten zeigen.



Es ist anzunehmen, dass sich bei fehlender Leitlinienempfehlung Unterschiede im Behandlungsverhalten zeigen.

Fallvignettenstudie

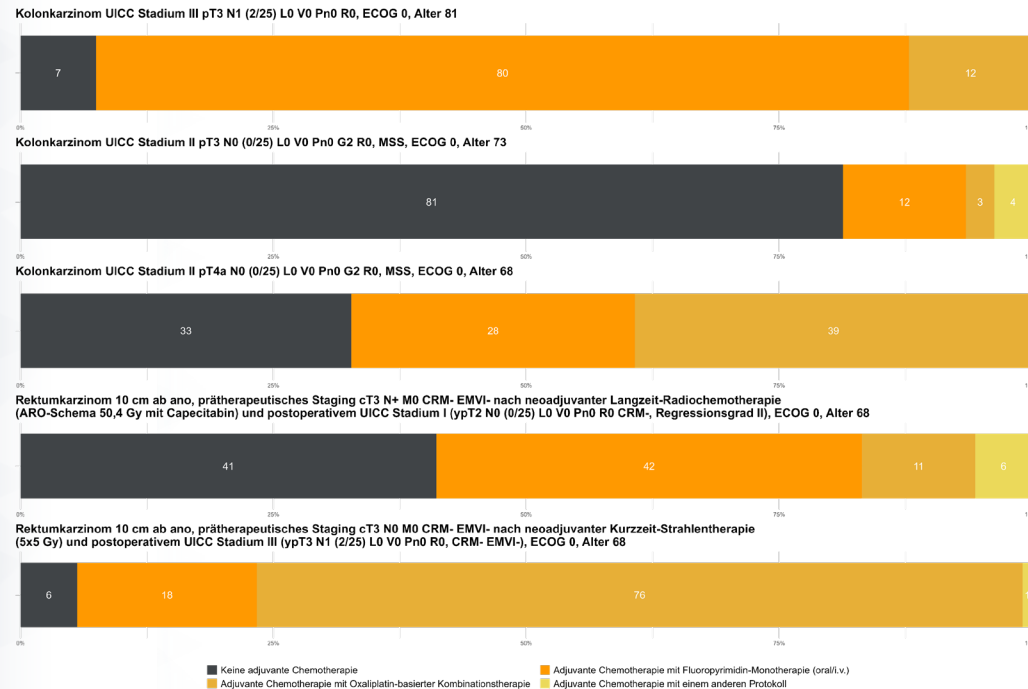
- **Anonyme Befragung** der Tumorboards von DKG-zertifizierten Darmkrebszentren zur **institutionsspezifischen Behandlungspräferenz**
- **Von 313** angeschriebenen Zentren wurden **117 Fragebögen zurückgeschickt. → Rücklaufquote von 37%**

Fragestellung

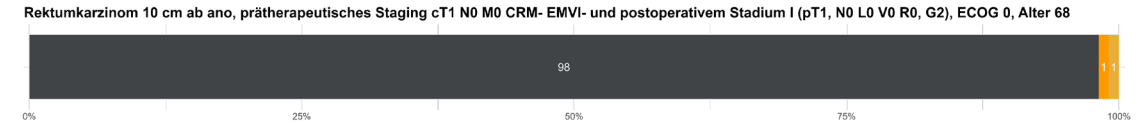
Fallvignettenstudie - Ergebnisse

Ohne Leitlinienempfehlung Ja, es ist heterogen!

N = 107, Häufigkeitsangaben in Prozent



Mit Leitlinienempfehlung Aber nicht in den Kontrollfällen



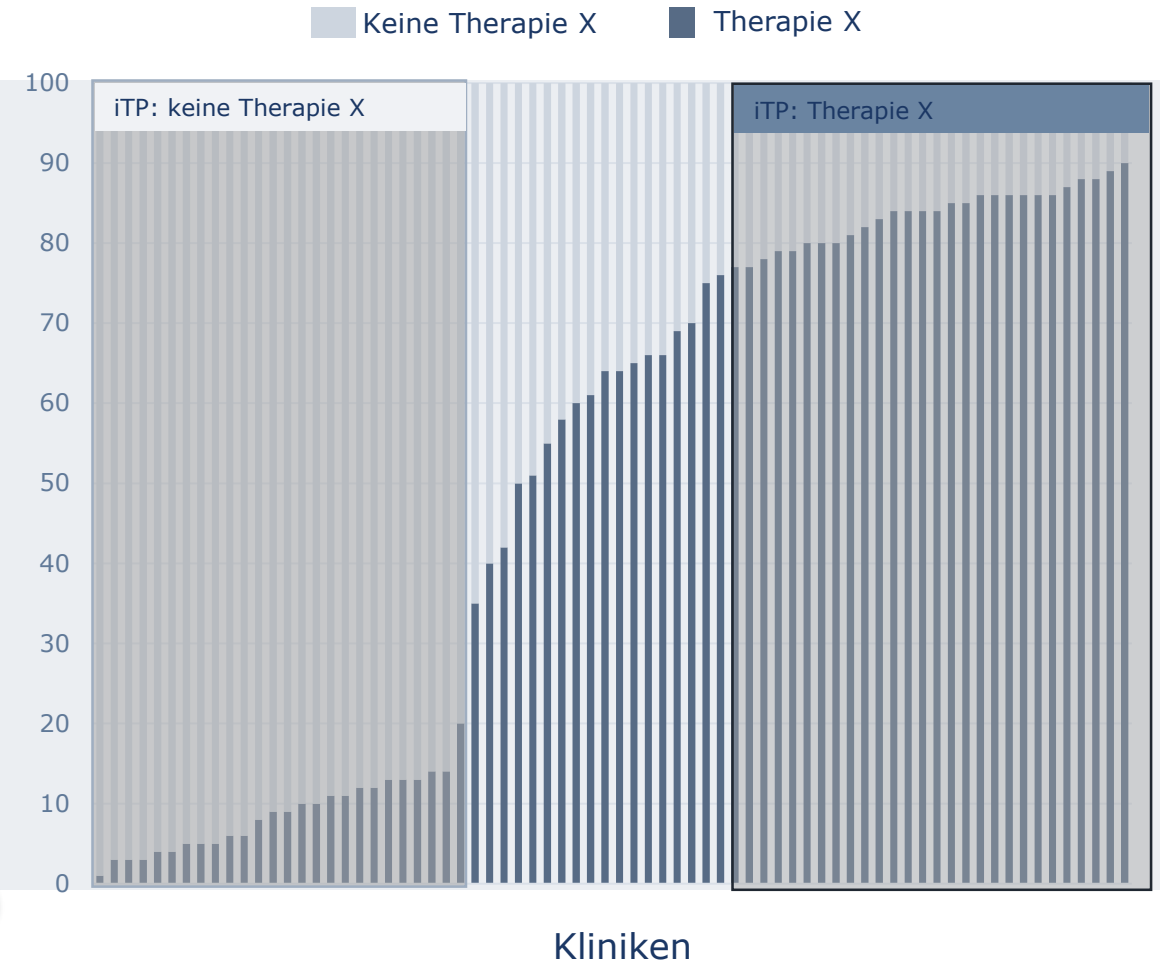
- Die präferenzbasierte Auswertung nimmt an, dass Unterschiede im **Behandlungsverhalten** zwischen **Institutionen** tatsächlich Unterschiede in der **Präferenz** widerspiegeln.

Institutionelle Therapiepräferenz

Fragestellung

Institutionelle Therapiepräferenz
in der AG KI erarbeitete Definition:

Die institutionelle Therapiepräferenz (iTP) ist die vom medizinischen Fachpersonal einer Klinik **empfohlene Therapie** (spezifische therapeutische Intervention/en) für einen klinisch definierten **Fall** (bestehend aus Diagnose, Stadium/Schweregrad, Alter, Geschlecht, Komorbiditäten) in dieser **Klinik**.



Projektziele

übergeordnetes Ziel:

- **Proof of concept** einer Methodik für eine IT-Infrastruktur zur Zusammenführung der Datensammlungen unterschiedlicher Datenquellen

Teilziele:

- Datenauswertung mit KI im Projekt und **konventionellen statistischen Methoden** → Abgleich der Ergebnisse
- Abbildung von **Institutionelle Therapiepräferenz**
→ Möglichkeit zur Reduktion von Confounding
- Generierung hochwertiger Evidenz für therapierrelevante Fragestellungen

KI im Projekt

Übersicht KI-Modelle

- **KI (Künstliche Intelligenz)**

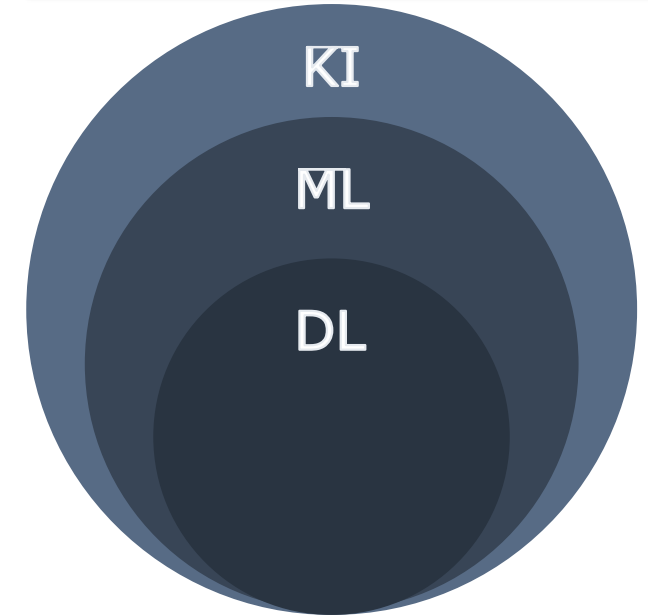
Systeme, die Aufgaben übernehmen können, die normalerweise menschliches Denken erfordern, wie zum Beispiel Probleme lösen, Entscheidungen treffen oder Sprache verstehen.

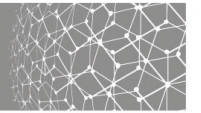
- **ML (Maschinelles Lernen)**

KI Modell bei dem das System aus vorhandenen Daten lernt. Sie können Muster erkennen und diese nutzen, um Vorhersagen zu treffen oder neue Entscheidungen zu fällen.

- **DL (Deep Learning)**

Deep Learning ist ein spezieller Bereich des maschinellen Lernens. Es verwendet künstliche neuronale Netze, die aus mehreren Schichten bestehen, um sehr komplexe Aufgaben zu lösen, wie Bilder analysieren oder Sprache verstehen.





- **ML-Modell-Typ:**
Überwachtes Lernen (supervised learning).
- **Eigenschaften des ML-Modells:**
 - Einmalige Einrichtung, kein selbstständiges Weiterlernen.
 - Kein Zugriff auf weitere Daten oder das Internet.
 - Vergleichbar mit klassischen Datenverarbeitungsprogrammen.
- **Voraussetzungen zum trainieren eines ML-Modells:**
 - Größtmögliches Maß an Variablen/Informationen notwendig für die Stabilität der Ki
- **Vorteile durch Datenkombination**
 - Erhöhung der Datenqualität und -quantität durch Verknüpfung verschiedener Datenquellen
 - Verbesserung der Modellgeneralität und -genauigkeit





Klinische Krebsregisterdaten:

Erfassen klinische Merkmale und Behandlungsverläufe von Krebspatienten.



Abrechnungsdaten der Gesetzlichen Krankenkassen:

Bieten Informationen über weitere Diagnosen, durchgeführte Behandlungen und Therapien, und Komorbiditäten auch über die Onkologie hinaus.



Daten der DKTK-Standorte:

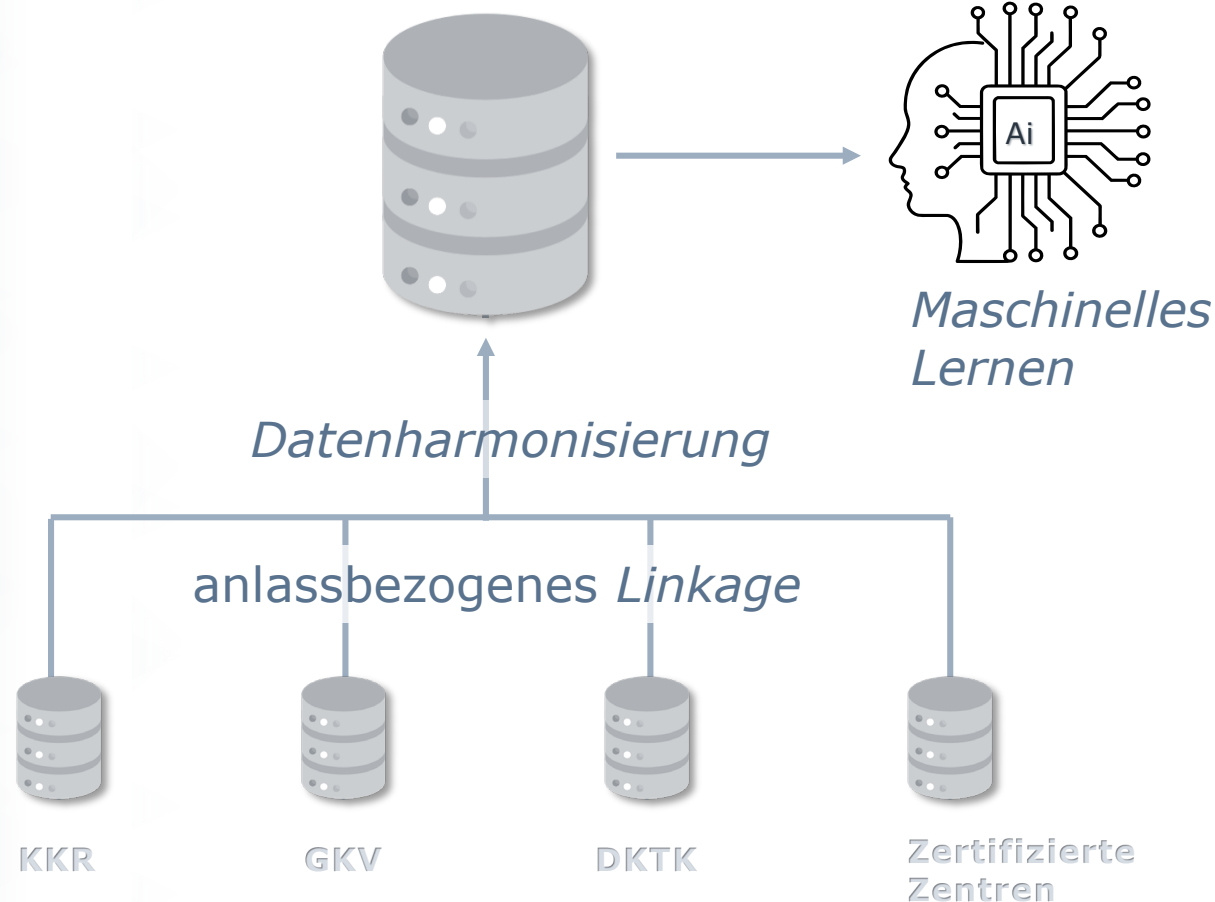
Stellen zusätzliche Daten zu klinischen Studien und Forschungsinitiativen bereit.



Daten der zertifizierten Zentren über OncoBox:

Spezifische Daten zu Onkologiebehandlungen und deren Resultaten.

Methodische Umsetzung Linkage



Methoden:

- *anlassbezogenes Linkage* von Daten aus verschiedenen Quellen (KKR, GKV, DKTK, Oncobox Research)
- umfassende *Datenharmonisierung*
- Nutzung moderner Verfahren des *maschinellen Lernens* und der Statistik

Methodische Umsetzung Linkage

SMPC- PPRL anlassbezogenes *Linkage*

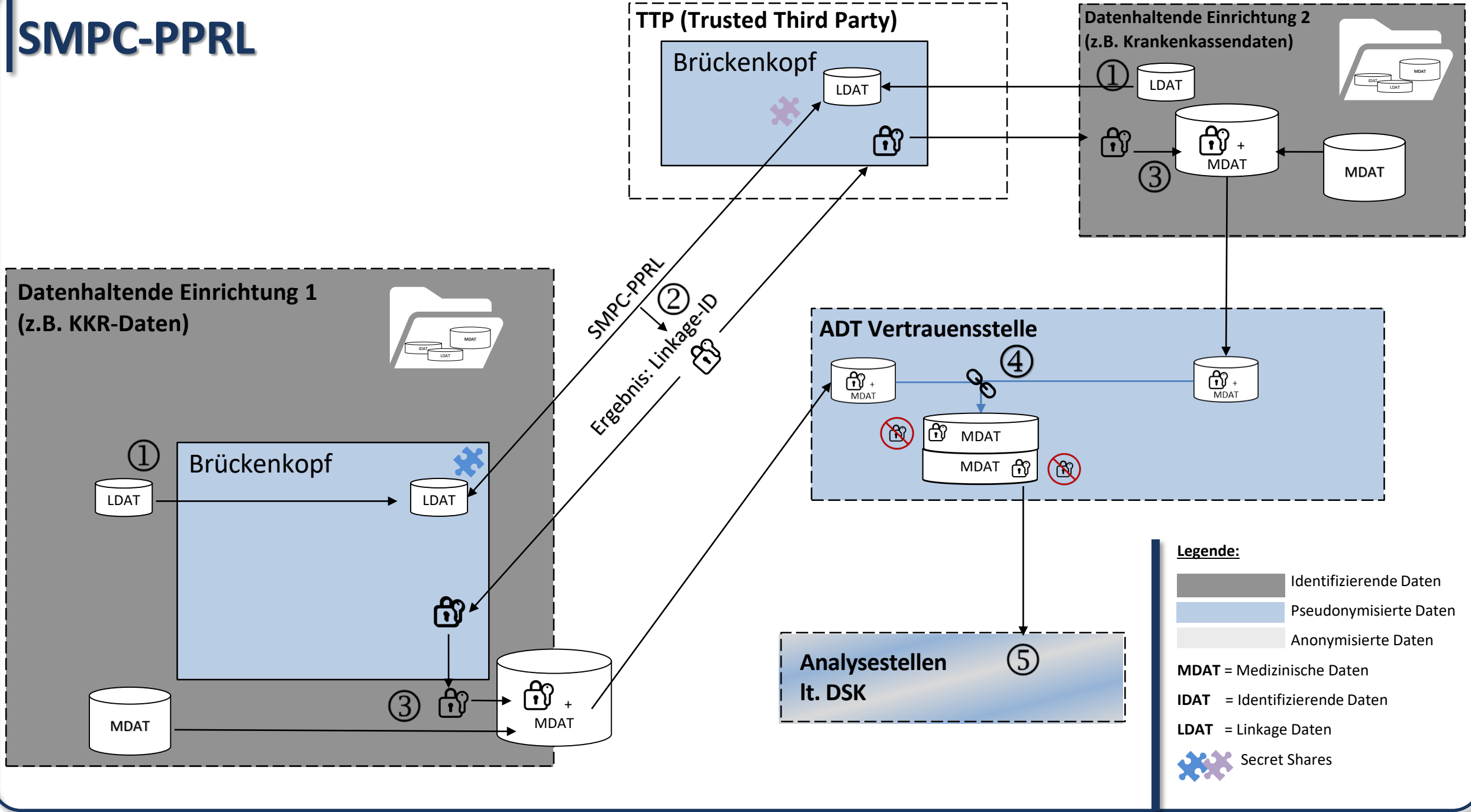
Secure Multi-Party Computation Privacy - Preserving Record

Mehrere Parteien führen Berechnungen gemeinsam durch, ohne die Daten der anderen zu kennen

Ziele von SMPC PPRL:

- Verknüpfung von Datensätzen aus verschiedenen Quellen, ohne die Privatsphäre zu gefährden
- Sichere und vertrauliche Datenverarbeitung über mehrere Parteien hinweg
- Keine direkte Weitergabe von vollständigen - identifizierenden Datensätzen für ein sicheres Linkage

SMPC-PPRL



Was haben wir bisher erreicht?

Klinische Expertise:

Konsultationen durchgeführt und dokumentiert

Studienprotokoll:

Studienprotokoll erstellt

Statistischer Analyseplan:

Abgeschlossen

Ethikvoten:

positive Ethikvoten in Regensburg, Dresden, Freiburg, Frankfurt eingeholt

Publikationen:

Fallvignettenstudie veröffentlicht

Testdatensätze:

Synthetische KKR-, GKV- und oncoBox-Daten erfolgreich erstellt

Datenlinkage, Transfer & Harmonisierung:

Datensatzdefinitionen festgelegt und abgestimmt → Konzept erstellt

Datenschutzkonzept:

Finalisiert und anwendbar

Pseudonymisierung & Transfer:

Konzept liegt vor, Testdatensätze erstellt

Pilotierung SMPC-PPRL:

Beweis der Funktionalität in hochsicheren Umgebungen wie Klinische Krebsregister

Feature Engineering (Testdaten):

Testdaten an AG KI übermittelt, Limitationen dokumentiert

Vernetzung, Erfahrungen und besondere Erfolge:

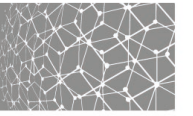
- Konsortium erfolgreich vernetzt: Verträge, Abstimmungen,
- Zusammenarbeit mit Kassen & Registern angestoßen
- Verfahren mit synthetischen Daten erprobt (Linkage, Pseudonymisierung, KI)
- Proof-of-Concepts zeigen Machbarkeit, erste Publikationen in Arbeit
- oncoFDZ Nachwuchs ;)

Beantragung der Daten:

- Übersicht der verschiedenen Antragsformen
- Zusage einer Datenlieferung von der Charité (DKTK Standort) und AOK Bayern.
- Absage vom Hessischen Krebsregister nach Einschätzung durch Datenschützer
- Weiter Rückmeldungen ausstehend

Wie geht es weiter?

- Kostenneutrale Verlängerung bis Februar 2026
- Rückmeldung Anträge auf Datennutzung
- Einschätzung der Linkage-Möglichkeiten verschiedener Datensätze
(Auslegungen von Gesetzesgrundlagen)
- Einschätzung bezüglich der Datenqualität und Datenquantität für die
Auswertung mittels Maschinellen Lernen.



Vielen Dank
Für Ihre Aufmerksamkeit