

# Synthese der Erfahrung aus den Use Cases

## Ein Überblick

Knut Kaulke

Workshop zum White Paper  
"Verbesserung des Record  
Linkage in Deutschland"

24. Mai 2022, Berlin

# Synthese der Erfahrung aus den Anwendungsfällen

- **Identifikatoren:** direkt basierend auf Identitätsdaten (IDAT wie z.B. Name, Vorname, Geburtsdatum, KVNR und Adresse)
  - Kodiert oder Klartext
- **Linkage-Verfahren**
  - Exaktes Record Linkage oft in lokalen klinischen Systemen (aber auch außerhalb z.B. Use Cases CoVerlauf und SHIP)
  - Mehrzahl der Use Cases im White Paper → probabilistisches Record Linkage
    - **Identifikatoren:**
      - Klartext oder
      - Kodiert und vergleichbar (Kontrollnummern oder Bloomfilter) → Privacy-Preserving Record Linkage (PPRL)

# Synthese der Erfahrung aus den Anwendungsfällen

- **Datenschutzkonforme organisatorische Umsetzung des Record Linkage: Informationelle Gewaltenteilung**
  - = Unterteilung der Forschungsinfrastrukturen in unabhängige Teilstrukturen/Module → z.B. Modul für Identitätsmanagement
  - Mehrzahl der Use Cases im White Paper → Einbindung einer zentralen, unabhängigen und vertrauenswürdigen Stelle (Treuhandstelle)
    - Separate Administrierung der personenbezogenen Daten von den für die Forschung relevanten medizinischen Daten → elektronisch geführte Patienten-/Probandenliste
    - Verzeichnis der Zuordnung von Patientenidentitäten (kodiert/Klartext-IDAT) zu Pseudonymen → enge Verzahnung von Verknüpfung der Identitäten und Pseudonymzuweisung
  - Föderiertes Record Linkage:
    1. Zusammenführung klinischer Daten auf lokaler Ebene
    2. Einrichtungsübergreifende Zusammenführung über kodierte IDAT (PPRL)

# Anwendungsfälle im White Paper

- Standortübergreifende Forschungsinfrastrukturen mit Primär- und Sekundärdaten
  - Use Cases: CoVerlauf, SHIP,...
  - Rechtliche Grundlage: Einwilligung
  - Modell: informationelle Gewaltenteilung
  - Umsetzung des Datenschutzes: mit oder ohne zentrale, rechtlich, räumlich und personell von der Forschung getrennten Treuhandstelle
  - Identifikatoren: kodierte oder unkodierte Identitätsdaten (IDAT)
  - Linkage-Verfahren: exaktes Linkage

# Anwendungsfälle im White Paper

- Standortübergreifendes Record Linkage mit Registerdaten
  - **Use Cases:** DFG-Linkage, Landeskrebsregister Nordrhein-Westfalen und Niedersachsen,...
  - **Rechtliche Grundlage:** §75 SGB X und länderspezifische Krebsregistergesetze
  - **Modell:** informationelle Gewaltenteilung
  - **Umsetzung des Datenschutzes:** mit (direkte Identifikatoren) oder ohne (indirekte Identifikatoren) zentraler, rechtlich, räumlich und personell von der Forschung getrennten Treuhandstelle
  - **Identifikatoren:** indirekte Identifikatoren oder kodierte IDAT
  - **Linkage-Verfahren:** probabilistisches Linkage oder deterministisches Linkage

# Anwendungsfälle im White Paper

- Standortübergreifende Forschungsinfrastrukturen mit klinischen Daten
  - Use Cases: nNGM
  - Rechtliche Grundlage: Einwilligung
  - Modell: informationelle Gewaltenteilung
  - Umsetzung des Datenschutzes: zentrale, rechtlich, räumlich und personell von der Forschung getrennten Treuhandstelle
  - Identifikatoren: unkodierte IDAT
  - Linkage-Verfahren: probabilistisches Linkage

# Anwendungsfälle im White Paper

- Standortinterne Verknüpfung klinischer Daten
  - **Use Cases:** insbesondere Universitätskliniken
  - **Rechtliche Grundlage:** Landeskrankenhausgesetze
  - **Modell:** eingeschränkte informationelle Gewaltenteilung = lokale Treuhänderservices
  - **Umsetzung des Datenschutzes:** zentrale, räumlich und personell von der Forschung getrennte Treuhandstelle
  - **Identifikatoren:** unkodierte IDAT
  - **Linkage-Verfahren:** exaktes Linkage

# Anwendungsfälle im White Paper

- Verteilte Erschließung von klinischen Sekundärdaten
  - Use Cases: DKTK, MII, NUM,...
  - Rechtliche Grundlage: Einwilligung und Landeskrankenhausgesetze
  - Modell: verteilte TTP-Services
  - Umsetzung des Datenschutzes: zentrale, rechtlich, räumlich und personell von der Forschung getrennten Treuhandstelle
  - Identifikatoren: kodierte IDAT
  - Linkage-Verfahren: probabilistisches Linkage



# Herausforderungen bei den Anwendungsfällen

- Technische Herausforderungen

- Gefahr von sowohl falsch positiven (Homonymfehler) als auch falsch negativen (Synonymfehler) Zuordnungen.
- Manuelle Überprüfung/Nachbearbeitung des Abgleichs → möglicherweise aufwändige händischer Vergleich der IDAT nötig
- Verfügbarkeit von KVNR
- Heterogenität/Harmonisierungsbedarf der Quelldatensätze
- Unterschiedliche technische Ausstattungen beim Datengeber

- Organisatorische/Rechtliche Herausforderungen:

- (Nachträgliche) Einwilligungsquote
- Unterschiedliche/spezifische rechtliche Regelungen zum Record Linkage bei z.B.
  1. Gesetzlichen und privaten Krankenkassendaten
  2. Landesgrenzendübergreifenden Verknüpfungen
- Aufwand für unterschiedliche Genehmigungsverfahren bei Nutzung der Daten mehrerer gesetzlicher Krankenkassen oder Krebsregistern aus mehreren Bundesländern

**Vielen Dank für die  
Aufmerksamkeit!**



**Kontakt:**

**Knut Kaulke**

[Knut.kaulke@tmf-ev.de](mailto:Knut.kaulke@tmf-ev.de)

**TMF e.V.**

[www.nfdi4health.de](http://www.nfdi4health.de)

TMF – Technologie- und Methodenplattform  
für die vernetzte medizinische Forschung e.V.



Gefördert durch



Deutsche  
Forschungsgemeinschaft

