

KIMONA – KI-basierte Meldungsoptimierung für Krebsregister in NRW

Louise Bloch¹, Chaimaa Karioui¹, Katarzyna Wojcik¹, Florian Oesterling¹, Viktor Pfaffenrot¹

¹ Landeskrebsregister NRW gGmbH, Gesundheitscampus 10, 44801 Bochum

HINTERGRUND

- Die qualitativ hochwertige und zeitnahe Krebsregistrierung bildet die Grundlage für Forschung unter anderem zur Versorgungs- und Pflegequalität von Krebserkrankungen
- Der onkologische Basisdatensatz (oBDS) ermöglicht die Erfassung des gesamten Behandlungsverlaufs
- Viele Felder des oBDS werden zur Zeit manuell aus in den Primärsystemen vorliegenden Freitexten (z.B. aus Arztbriefen oder Pathologieberichten) der Melder übertragen

IDEE

- Methoden der künstlichen Intelligenz (KI) können Informationen aus unstrukturierten Fließtexten extrahieren
- Zudem können KI-Methoden dabei helfen Meldeanlässe zuverlässiger zu erkennen
- Die automatisierte Informationsextraktion kann die Qualität und Aktualität der Krebsregistrierung verbessern sowie den manuellen Dokumentationsaufwand reduzieren

UMSETZUNG

- Die Comprehensive Cancer Center (CCC) der Unikliniken in Nordrhein-Westfalen (NRW) tragen Datenartefakte mit Extract, Transform, Load (ETL)-Prozessen in Data-Lake-Infrastrukturen zusammen und optimieren schließlich den Meldeprozess mit KI-Pipelines (Fig. 1A)
- Das Start-Up TamedAI unterstützt bei der Extrahierung der relevanten Informationen aus unstrukturierten Daten
- Das Landeskrebsregister NRW (LKR NRW) validiert unter anderem die Vollständigkeit, Vollständigkeit, Qualität und Vergütungsfähigkeit der KI-Meldungen und überprüft die Kompatibilität mit Folgeprozessen (epidemiologisches und klinisches Best-of (CBO)) (Fig. 1B)

- Zur Schätzung der erwartbaren Anzahl meldepflichtiger Ereignisse (v.a. Therapie- und Verlaufsmeldungen) werden statistische Modelle eingesetzt
- Anders als die Erwartungswertmodelle des Robert-Koch-Instituts (RKI), welche die Anzahl inzidenter Krebsfälle modellieren, werden unterschiedliche Meldeanlässe auf Einrichtungsebene geschätzt
- Die statistischen Modelle verwenden zusätzliches Wissen, z.B. S3-Leitlinien oder Qualitätsberichte
- Das zusätzliche Wissen soll unter anderem mittels KI-Modellen verfügbar gemacht werden
- Aus Datenschutzgründen werden Open-Source bzw. Open-Weights KI-Modelle auf lokaler Hardware (2 x NVIDIA H200 (PCIe 5.0 x16, 141 GB HBM3e je GPU)) eingesetzt

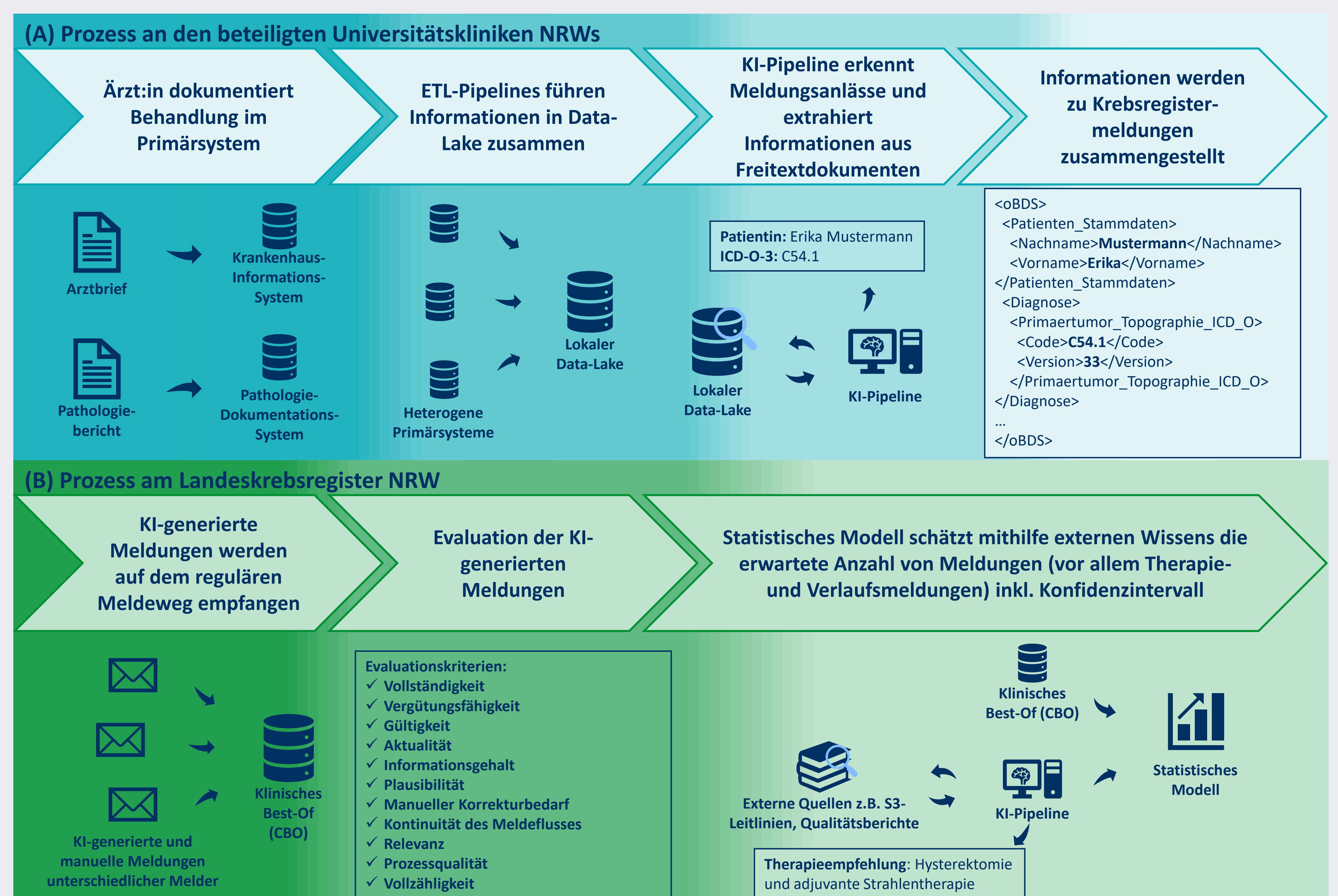


Fig. 1 Schematische Übersicht über die im Projekt geplanten Prozesse der Unikliniken (A) und des LKR NRW (B).

AKTUELLER STAND

- Key-Performance-Indicators (KPIs) für die Evaluation der KI-generierten Meldungen sind definiert
- Aktuell werden weitere Quellen identifiziert, die Hinweise auf fehlende Meldungen für die Schätzung der erwarteten Meldungsanzahl geben
- Zudem werden relevante statistische Modelle zur Schätzung der erwarteten Anzahl von Therapie- und Verlaufsmeldungen recherchiert
- Als nächstes sollen die identifizierten Quellen mithilfe einer KI-Pipeline, welche ein Retrieval-Augmented-Generation (RAG)-System enthält, verfügbar gemacht werden