

Datenverarbeitungspipeline zum ZfKD-Lieferdatensatzformat

Ole Johanns – Hamburgisches Krebsregister



27.09.2023

Informationstagung
Tumordokumentation
und Krebsregistrierung



Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH



UNIVERSITÄT ZU LÜBECK
INSTITUT FÜR SOZIALMEDIZIN
UND EPIDEMIOLOGIE



Krebsregister
Saarland ●●●●



ZENTRUM FÜR
KREBSREGISTERDATEN



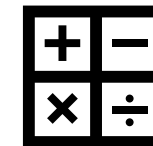
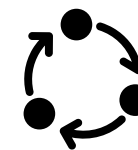
Ausgangslage

*KI-unterstützte, versorgungsnaher Nutzung von
Krebsregisterdaten*

Ausgangslage

- Datenerhebung durch onkologischen Basisdatensatz
- Registerspezifische und dadurch nicht einheitliche Datenverarbeitung
- Klinischer ZfKD-Lieferdatensatz als einheitliches Datenformat (XML)

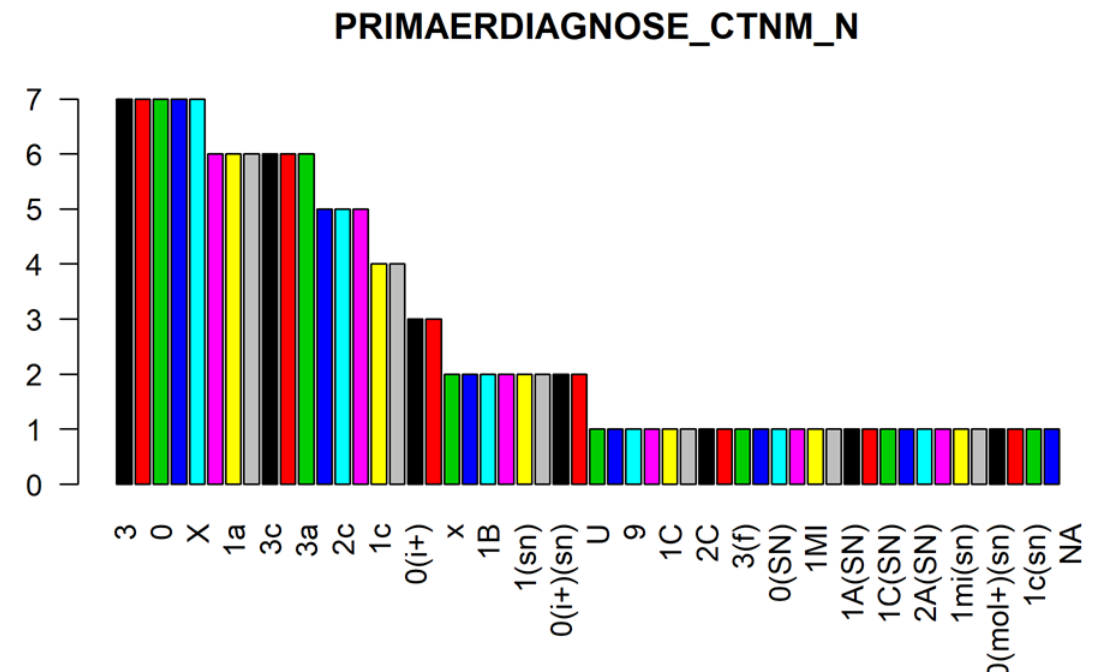
**BUNDESEINHEITLICHER
ONKOLOGISCHER
BASISDATENSATZ**



```
Schema_Version="3.0.0.8_RKI" xsi:schemaLocation="http://www.basisdatensat
<Lieferregister Register_ID="12"/>
<Lieferdatum Datumsgenauigkeit="E">2019-04-05</Lieferdatum>
▼<Menge_Patient>
  ▼<Patient Patient_ID="467828">
    ▼<Patienten_Stammdaten>
      <Geschlecht>W</Geschlecht>
      <Geburtsdatum Datumsgenauigkeit="M">1983-04-08</Geburtsdatum>
    ▼<Vitalstatus>
      <Verstorben>J</Verstorben>
      ▼<Todesursachen>
        ▼<Menge_Weitere_Todesursachen>
```

Welche Ausprägungen haben (Freitext-)Variablen?

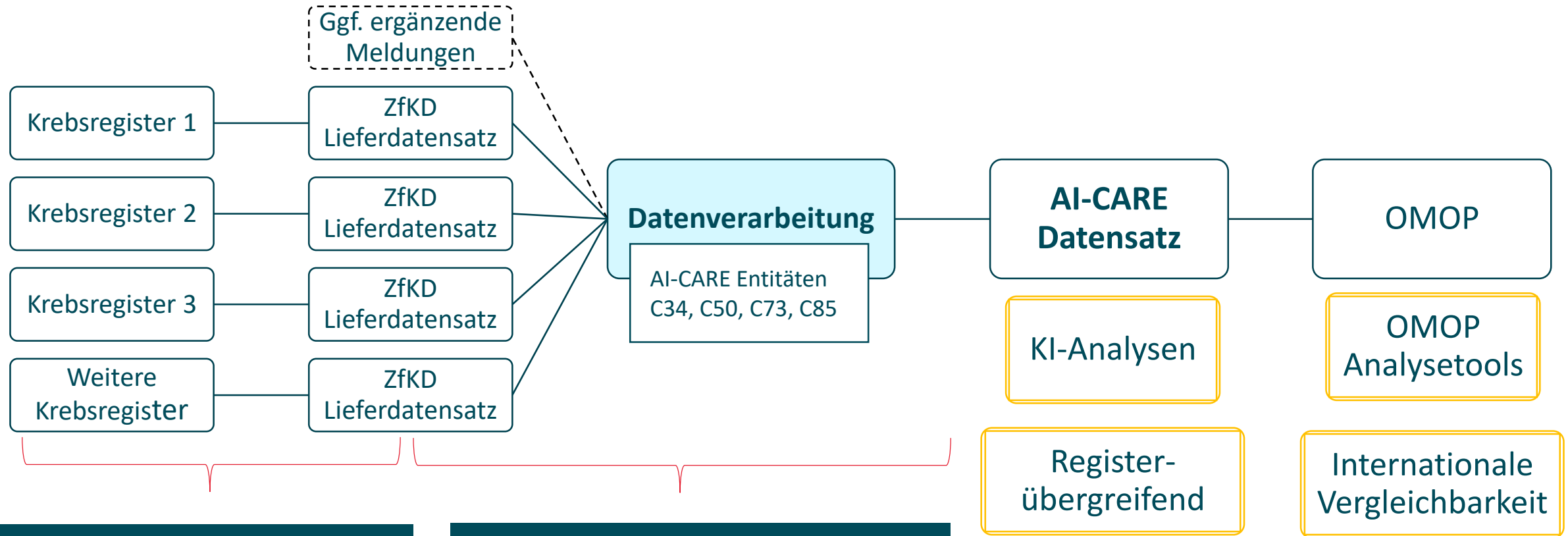
- **Dezentral:** Datenprofilierung mit WhiteRabbit durchgeführt, Open Source Werkzeug der OHDSI ([WhiteRabbit for ETL design – OHDSI](#))
- **Zentral:** Übersicht der Ausprägungen von 40 oBDS Variablen in den teilnehmenden Registern erzeugt.
- **Nutzen:** Mit welchen (Freitext-)Angaben müssen wir im Projekt rechnen? Welche Vereinheitlichung bringt die ZfKD-XML?



AP2: Datenaufbereitung für Verfahren der Künstlichen Intelligenz

Datenverarbeitungspipeline

AP2: Datenaufbereitung für Verfahren der Künstlichen Intelligenz



Unterschiede zwischen den Krebsregistern
 Vortrag Vera Heinrichs

Datenverarbeitung
 Vereinheitlichung der Daten
 Datenqualitätsanalysen

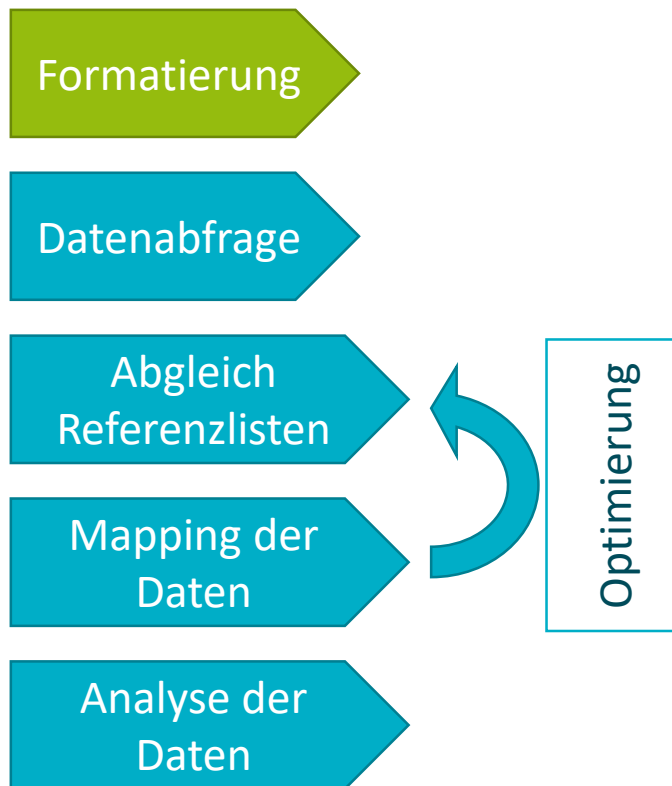
ZfKD - Lieferdatensatz

Variable	Gehört zu	oBDS Modul
72 Ulzeration	Primärdiagno	Modul Melanor
73 LDH	Primärdiagno	Modul Melanor
74 Rektum: Abstand des Tumorunterrandes zur Anokut	Primärdiagno	Modul Darm
75 Mutation K-ras-Onkogen	Primärdiagno	Modul Darm
76 Operationsdatum	Operation	13.2
77 Art der Operation nach OPS Code	Operation	13.3
78 Art der Operation nach OPS Version	Operation	13.4
79 Intention der OP	Operation	13.1
80 Lokale Beurteilung Residualstatus	Operation	10.1
81 Abstand Primärdiagnose bis Operation	Operation	neu (berechnet)
82 Beginn Strahlentherapie	Strahlenther	14.5
83 Intention	Strahlenther	14.1
84 Stellung zur OP	Strahlenther	14.2
85 Applikationsart	Strahlenther	14.7
86 Zielgebiet	Strahlenther	14.3
87 Seite Zielgebiet	Strahlenther	14.4
88 Dauer Strahlentherapie	Strahlenther	neu (berechnet)
89 Abstand Primärdiagnose bis Beginn Strahlentherapie	Strahlenther	neu (berechnet)
90 Intention der systemischen Therapie	systemische	16.1
91 Stellung systemische Therapie zur OP	systemische	16.2
92 Art d. systemischen Therapie	systemische	16.3
93 Substanz	systemische	16.6
94 Protokoll	systemische	16.4
95 Beginn systemische Therapie	systemische	16.5

Anzahl der Variablen	Kategorie
8	Person
50	Primärdiagnose
17	Primärdiagnose Module
6	Operation
8	Strahlentherapie
8	Systemische Therapie
24	Folgeereignis

<https://plattform65c.atlassian.net/wiki/spaces/P6/pages/57507841/Download>

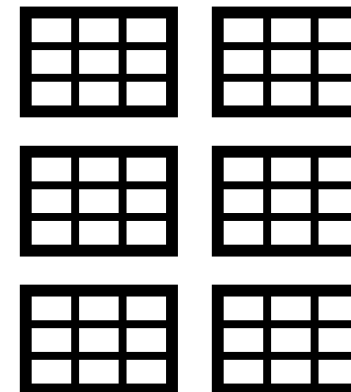
Formatierung



```

Schema_Version="3.0.0.8_RKI" xsi:schem
<Lieferregister Register_ID="12"/>
<Lieferdatum Datumsgenauigkeit="E">20
▼<Menge_Patient>
  ▼<Patient Patient_ID="467828">
    ▼<Patienten_Stammdaten>
      <Geschlecht>W</Geschlecht>
      <Geburtsdatum Datumsgenauigkeit
    ▼<Vitalstatus>
      <Verstorben>J</Verstorben>
    ▼<Todesursachen>
      ▼<Menge_Weitere_Todesursacher
  
```

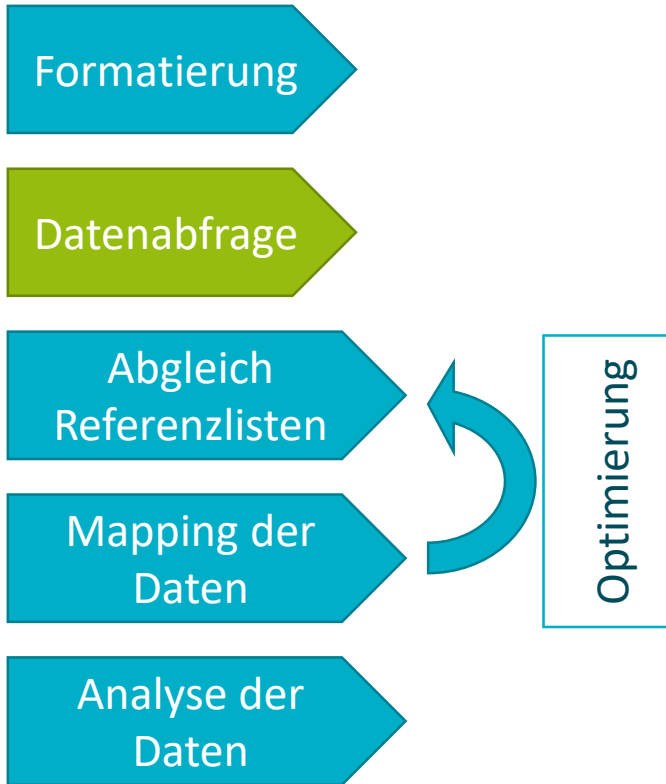
XML



CSV

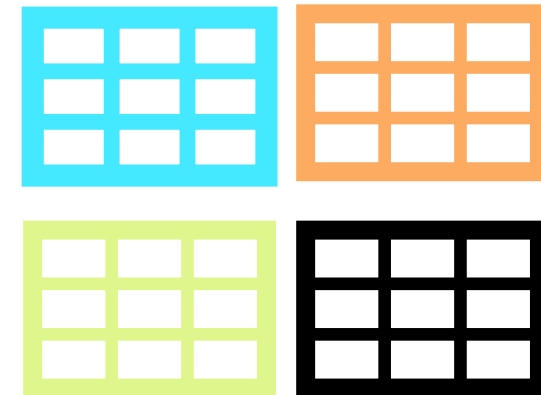
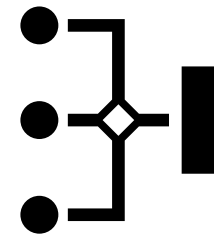
Umsetzung: Sebastian Germer, DFKI

Datenabfrage

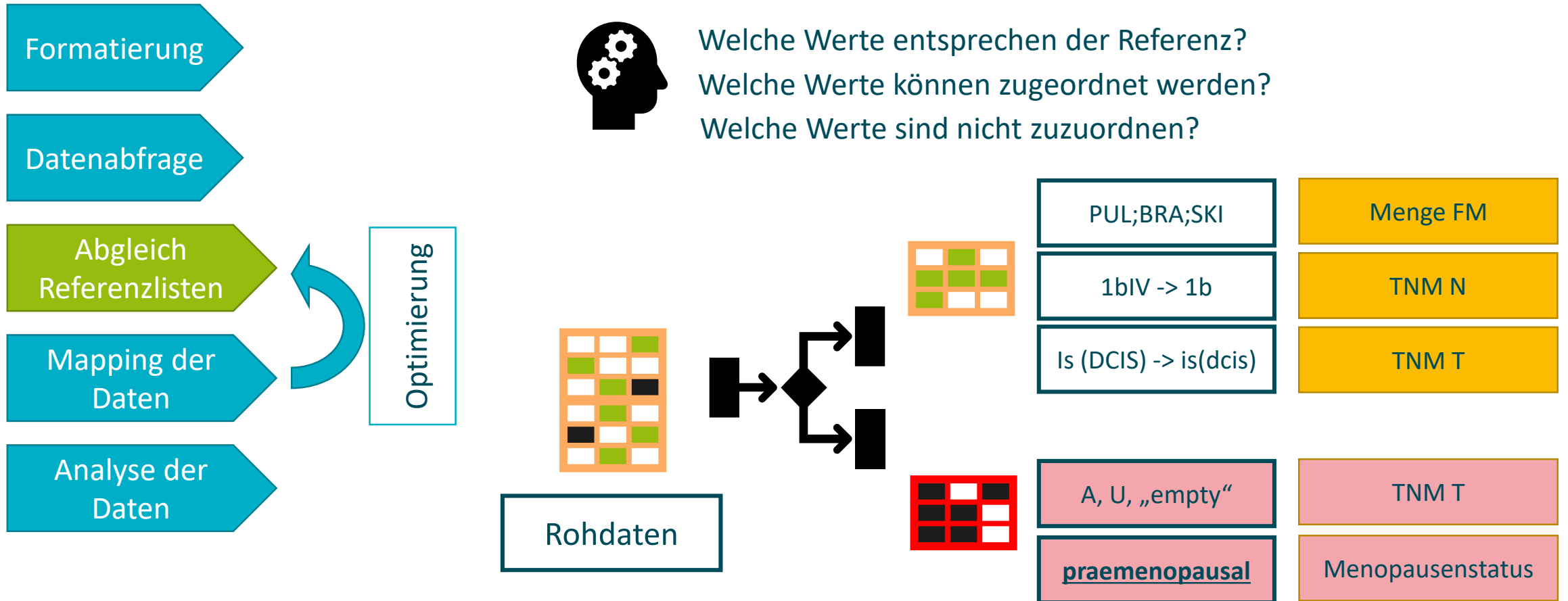


Welche Register liegen vor?
Welche Tabellen liegen vor?

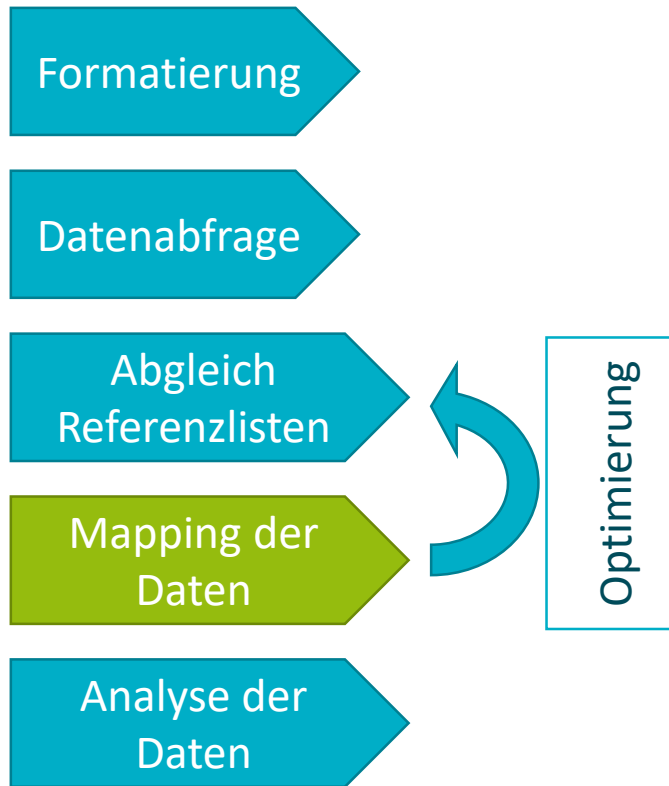
Abfrage aller
Tabellen eines
Registers



Abgleich mit Referenzlisten

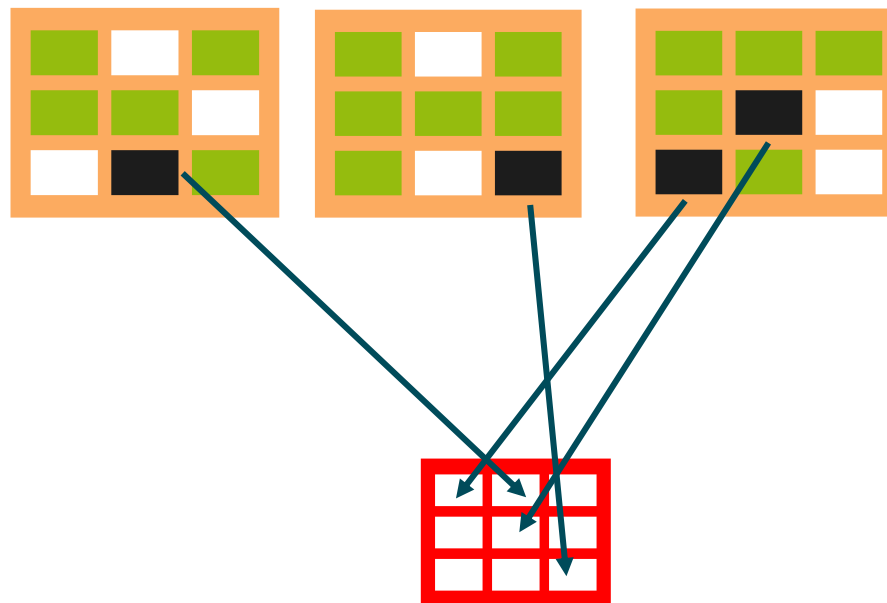


Mapping der Daten

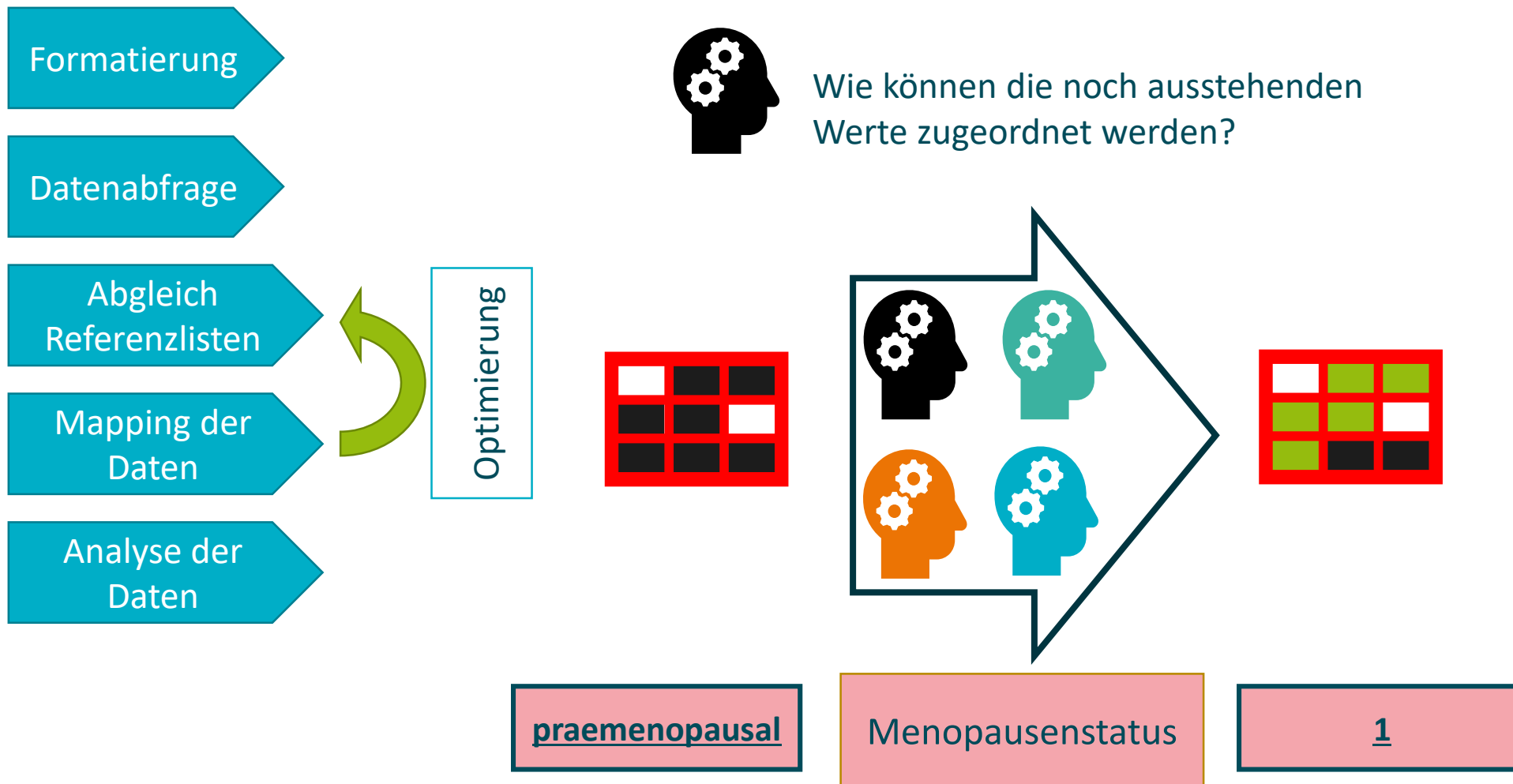


Korrektur der Ausgangswerte jeder Variable für jedes Register

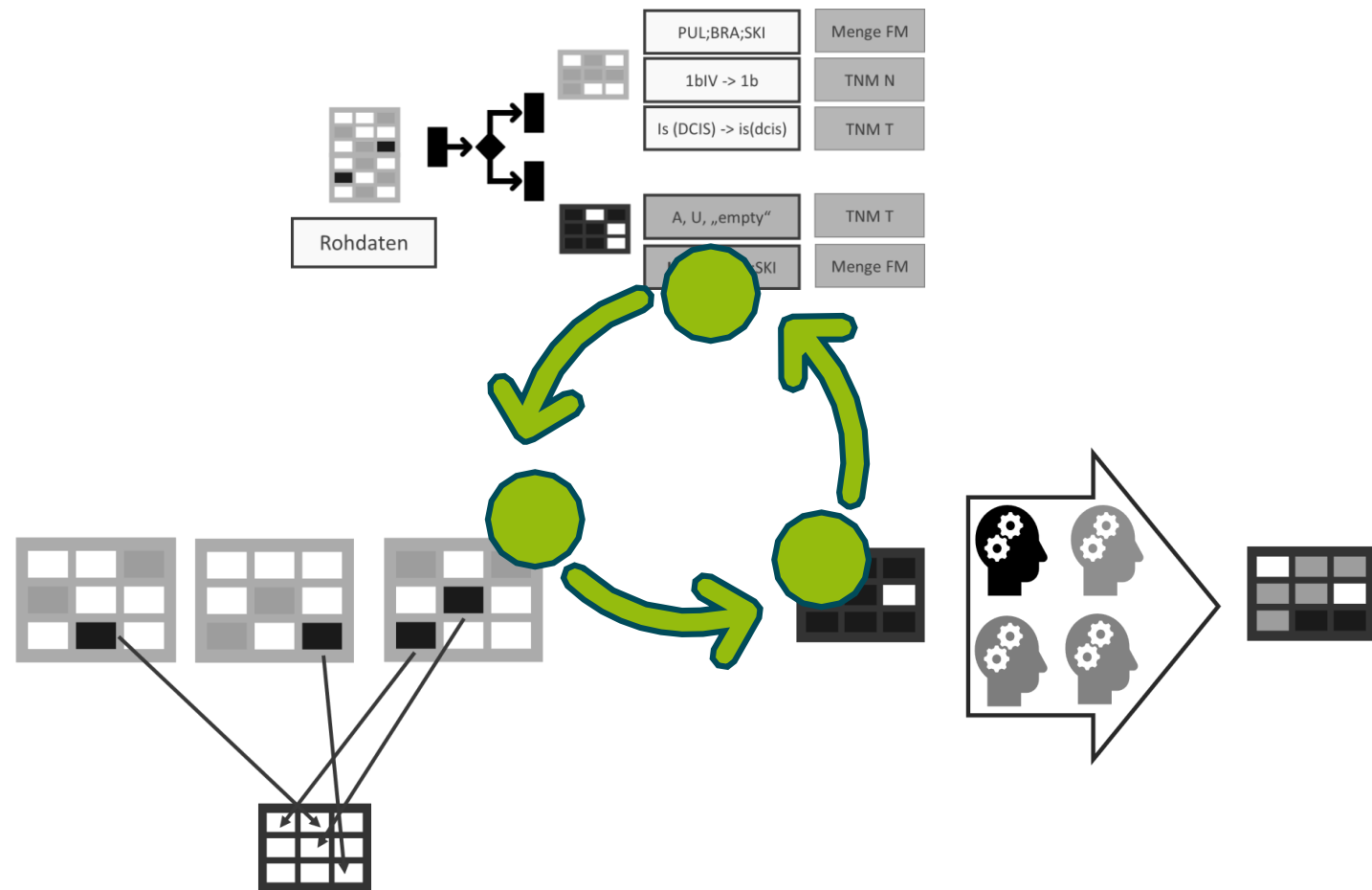
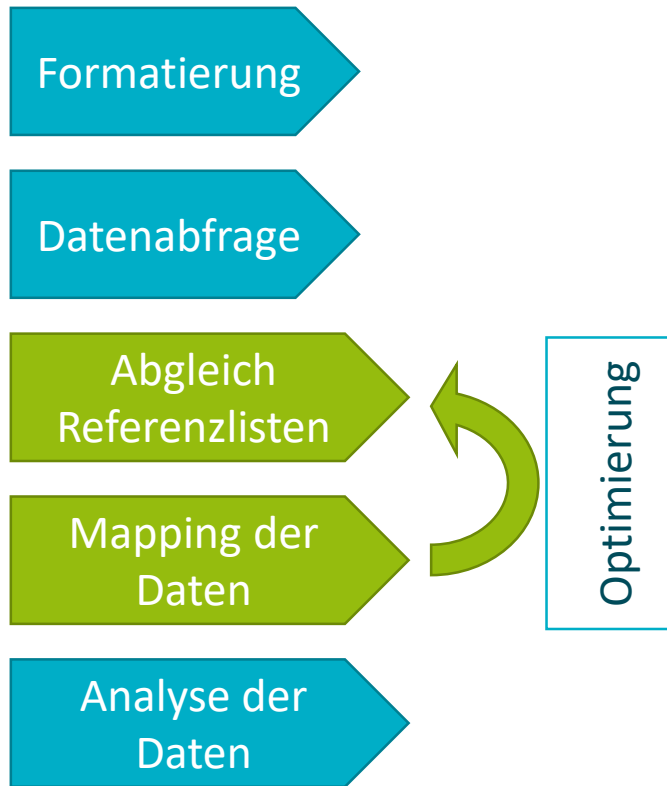
Sammlung der nicht zuzuordnenden Werte aller Register



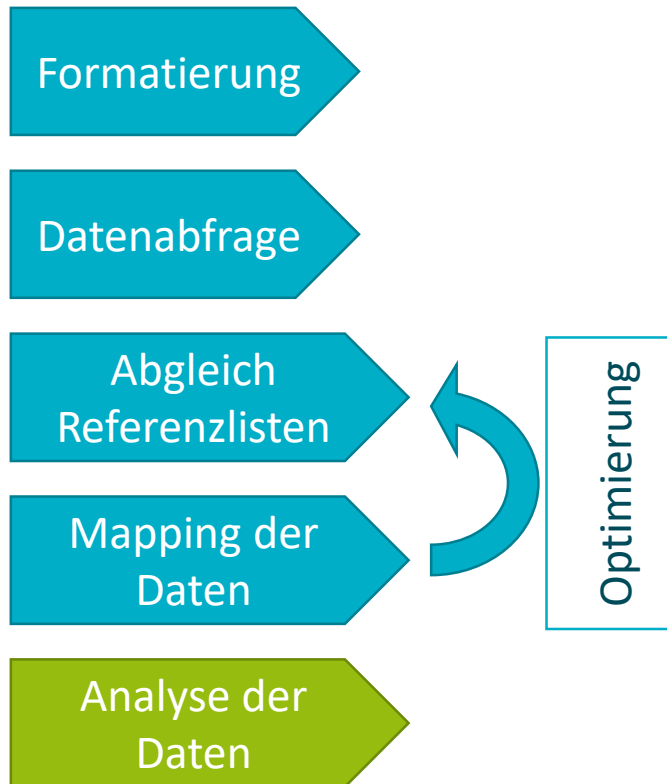
Verbesserung der Referenzlisten



Verbesserung der Referenzlisten

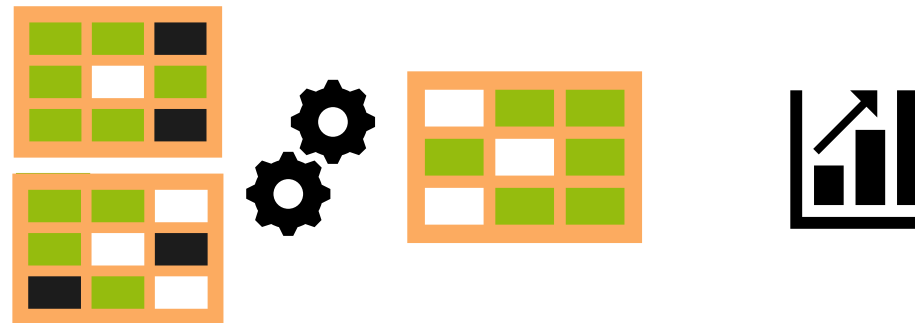


Analyse der Daten



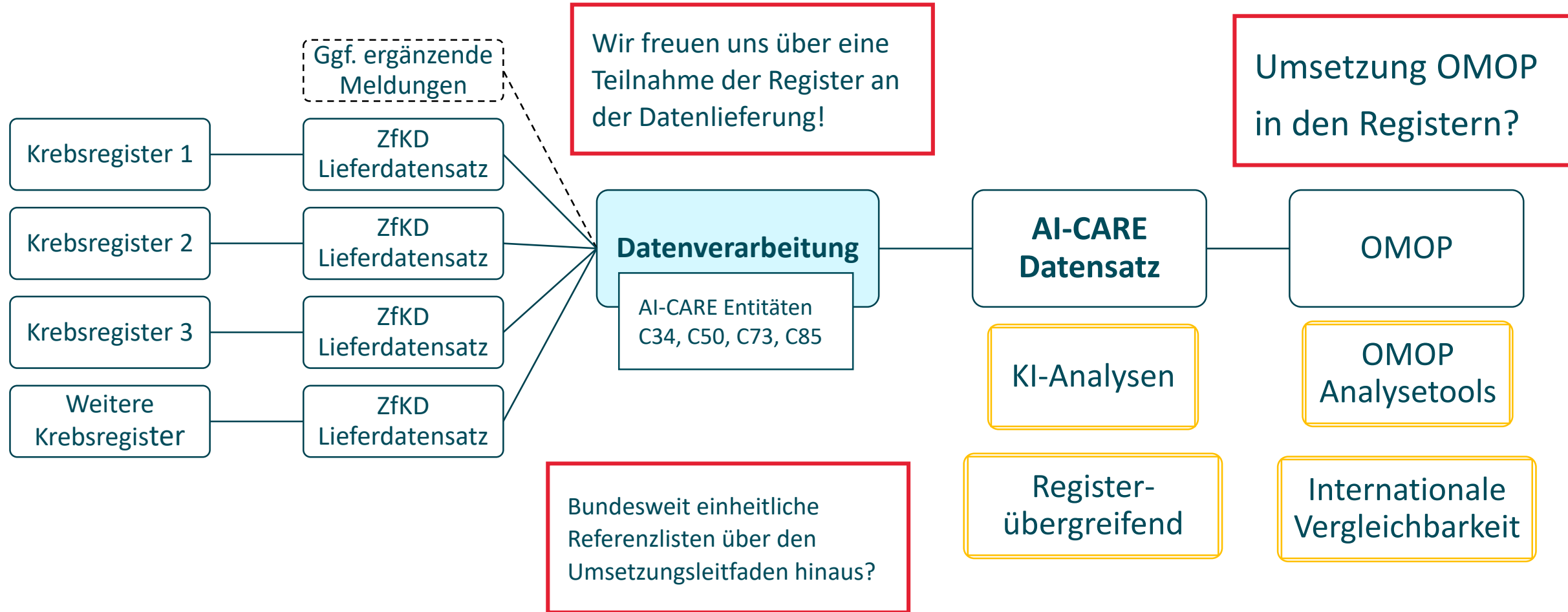
Analyse der Daten hinsichtlich Verteilung vor und nach der Datenverarbeitung

Wie profitieren die weiteren Analysen des Projektes von der Aufarbeitung und Zusammenstellung der Daten?



	Anteil nicht definierter Werte	Anteil fehlender Werte	Anteil und Anzahl auswertbarer Fälle
Nach Verarbeitung	sinkt	steigt leicht	steigt

AP2: Datenaufbereitung für Verfahren der Künstlichen Intelligenz



Vielen Dank für die Beteiligung

Sebastian Germer, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Hanna Hees, Universitätsklinikum Hamburg-Eppendorf

Vera Heinrichs, Hamburgisches Krebsregister

Alice Nennecke, Hamburgisches Krebsregister

Henrik Kusche, Hamburgisches Krebsregister

AI-CARE Working Group, www.ai-care-cancer.de

ENDE

